

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 07:40:07 ; Search time 9801 Seconds
(without alignments)
4454.459 Million cell updates/sec

Title: US-09-920-705-3

Perfect score: 4801

Sequence: 1 MALSMKPRADSDGFGSNL.....LAIQKQKQSSGTPAGCP 901

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh
-Q/cg2.1/USPTO.spool.h/US09920705/runat_07042005_101813_20638/app_query.fasta_1.1095
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09920705_@CGN_1_1_5932_@runat_07042005_101813_20638 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4801	100.0	2706	6	AX420420 Sequence
2	4786	99.7	2706	6	BT003029 Arabidops
3	4786	99.7	3498	8	AY094463 Arabidops
4	4656.5	97.0	3171	6	CQ804588 Sequence

ALIGNMENTS

RESULT 1	AX420420	Sequence 2 from Patent WO0212518.	2706 bp	DNA	linear	PAT 18-JUN-2002
LOCUS	AX420420					
DEFINITION	AX420420					
ACCESSION	AX420420.1	GI:21524576				
VERSION	AX420420.1	GI:21524576				
KEYWORDS						
SOURCE	Arabidopsis thaliana	(thale cress)				
ORGANISM	Arabidopsis thaliana					
REFERENCE	Amasino, R.M., Schomburg, F.M., Michaels, S.D. and Patton, D.					
AUTHORS	Patent: WO 0212518-A 2 14-FEB-2002;					
TITLE	WISCONSIN ALUMNI RESEARCH FOUNDATION (US)					
JOURNAL	Location/Qualifiers					
FEATURES	1..2706					
source	1..2706					
CDS	1..2706					
	/note="unannoted protein product"					
	/codon_start=1					
	/protein_id="CAD35485.1"					
	/db_xref="GI:21524577"					
	/translation="MALSMKPRADSDGFGSNLWVGSJLPETTESDITELFGRYGDI					

DRITVSSRGPAFTYRHHVEBAVAKALQGANLNGSIOKIEVAPAKPKSLMWGSI
 GPNVSKDLPEEFSEFKIEDFRFLRBRKTAFLIDYEMDALQAKSNMGKMSFLR
 VDLPSRQAPKKEOMAGSYDNNGNMNHRPOVPHYEPFKEDVPSKYLMPETATO
 CNDROILLANMILFGEIHERVSYPSRNPALVERFSAEAOCEKGLQRLFNPRKX
 MYSNDELPEDODTFSYSGMRSTDMFNNDPSCVSSPHSTGIPGSRPLRGTEBSY
 NGARINDVQKPEPMWRPSPANGTILPSPTGILPSPAGOTRRPMKSNDSWEGDP
 AQLVRESKTRRSDVDGFTPMGVDSRFGSGVAARPIGSPRSDIIMGMIAKGGT
 PVCCARCPWPKGILETKLPEVANCASRTDNLMAKXVAVALIGERIVFVREDPDS
 YTEFLRILYSKORBAVAKLDGDTLLFLIPPSDITLDTLOVTRBERLYGVVLLKPPAV
 PVTSYRQESQSNPLHYMDQADSDPANAHSILYPRRNTIRGAPHEHILASKESVSE
 LRIRPMNAPPOAGSVLTPELLATLASLILPATISQPAEPHSPMSGSTVSTAGSEL
 YNGAPSOQAMRGQIVADHANSOSFOQYGNQYPAQGLPPRPSRYPASNPMTYGM
 VHGMMQYOSQSVNMPOLSPENPHNNYSMTQSSNHPVPSQPMVOQYCPBAEMPON
 YGPIPSYQOANFHVITNOAONLPSQFOAMOPADKXALBEPONALRLQPMISDGD
 QGTTGDEVDKNORYOSTLQFANLLIQKQKQSSGTRPAGGP"
 46.279
 /bound_molecy="RNA Binding Region"
 283.522
 /bound_molecy="RNA Binding Region"
 610.852
 /bound_molecy="RNA Binding Region"

ALIGNMENT Scores:

Pred. No.: 9.13e-184 Length: 2706
 Score: 4801.00 Matches: 901
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-920-705-3 (1-901) x AX420420 (1-2706)

QY 1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
 DB 1 ATGGCGTTATCTAATGAAGCATTCAAGCCGATTCCTCGGTTTCAATCAATCAATCTT 60
 QY 21 TTPVALGlySerLeuThrProGluThrThrgLysSerAspLeuThrGluLeuPheGlyArg 40
 DB 61 TGGGTGCGTACCTTAACGCGGAGACGACAGACTGACATCTGACGAGTTGTTGGAAGA 120
 QY 41 TYGlyValAspIleAspArgIleThrValTYrSerSerArgGlyPheAlaPheIleTYr 60
 DB 121 TACGGCGATTTATTAATGATCAAGGTGATCTTCAAGAGCTTGGCTTTATATATAC 180
 QY 61 ArgHisValGluGluValAlaValAlaValysGluValAlaLeuGlnGlyValAsnLeuSngly 80
 DB 181 AGACATGTGGAGGAGACAGTCCGACGCAAGAGGCTCTTCAAGAGCAATTTGATGGA 240
 QY 81 SerGlnIleLysIleGluTYrAlaArgProAlaLysProCysLysSerLeuTTPVALGly 100
 DB 241 AGTCAATTAAGTTCGAATACGACACCGCAAACTTGTATAGTCTTATGGTGGGT 300
 QY 101 GLYIleGlyProAsnValSerLysAspAspLeuGluGluPheSerLysPheGlyLys 120
 DB 301 GGAATGGGCGCTTAATCTTCCAAAGATGACTCTGGAGAGAGAGTTCCGCAAGTTTGGGAAA 360
 QY 121 ILeGluAspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTYrTYrGluMet 140
 DB 361 ATTCAGAGATTTAGGTTCTTCAGAGAACGCAAGACAGCTTTCATTGATTAATATAGATG 420
 QY 141 AspAspAlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArg 160
 DB 421 GATGATGCTTTACAGGCTTAAGACATGATGAGAAAGCTTATGGGTGATGCTTTTGGGT 480
 QY 161 ValAspPheLeuArgSerGlnAlaProLysLysGluGlnTTPVALGlySerTYrAspAsn 180
 DB 481 GTTGAATTTCTCCGGTCAACAAGCAAAAGAAACATGGCGTGGCTCTTACATATAC 540
 QY 181 ArgAsnGlyAsnMetAsnHisLysProGlnTYrProHisSerTYrGluAspPheLysGly 200
 DB 541 AGAAATGGCAATATGAAATCATTAACCGCAGTATCTCTCACTCATATGAGACTTTAAAGA 600

QY 201 AspValGlnProSerLysValLeuTTPIleGlyPheProProThrAlaThrGlnCysAsn 220
 DB 601 GATGTCCACCAAGTAAAGGTTCTTGAGATGGGTTCCCTCTCACTGATCAACAATGCAAT 660
 QY 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
 DB 661 GATGAGCAATTTCTGCACATCGATCGATCTTCTTGAGATGAGAGGGGTAAAGT 720
 QY 241 TYrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluValAArgGlnCys 260
 DB 721 TACCATCAAGAAATTTTGCACTTGTTGAGGTTTAAAGAGGGAGAGAGCTGCCAATGC 780
 QY 261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProAlaGlyIleLysIleMetTYrSerAsn 280
 DB 781 AAGGAAGGCTTACAGGGAGGTTATTAATATATCTTAAGAACAAATATATGACTCAAAAC 840
 QY 281 AspGluLeuProProGluGluAspAspThrSerPheTYrSerGlyMetLysArgSerArg 300
 DB 841 GATGATGTCCTCTCGACAGACAGATCTAGTTTTCCTGTTGATGAACGGTCAAG 900
 QY 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
 DB 901 ACAGATATGTTCAATATATGATCTTCATGTATCTTCTCTCATTTCTACTGGAATTCCT 960
 QY 321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTYrAsnGlyValaGluTYrAsn 340
 DB 961 GGGTTATAGAGCCCTCTCAGAGGTACGATAGAGGCTTCAATATATGTTGTCAGAAATACAT 1020
 QY 341 AspValValaGlyLysGluProAsnTTPArgArgProSerAlaAsnGlyThrglyIleLeu 360
 DB 1021 GACGTTGTTGGTAAAGAGCCAACTGGAGAGGCGCATCTGAATATGAATCGAAATCTC 1080
 QY 361 ProSerProThrArgLysProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMet 380
 DB 1081 CCATCTTCAACAGACCTCGAATCTCTCCATCTCTCACAAGGTACGAGGCGCCCTATG 1140
 QY 381 ArgSerAsnProAspSerTTPGluGlyTYrAspProAlaGlnLeuValAArgGluSerLys 400
 DB 1141 AGTCAAAACCCGATTTCTTGGAAGATATGATCTCTCAAGTTGTTGCAAGAAAGTAAA 1200
 QY 401 ArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValaAspGluArgSer 420
 DB 1201 CGAACCAAGAAAGATGATCAGTGCAGGTTTATCTCAATGGGTGTCATGAGAGGTCA 1260
 QY 421 PheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHisIle 440
 DB 1261 TTGGTGAAGGTTCAAGTTCGCTAGACCTATCGTGGCCCCCTGATTTCTGATACATA 1320
 QY 441 TTPArgGlyMetIleAlaLysGlyGlyThrProValCysCysAlaArgCysValaProMet 460
 DB 1321 TGAGAGGAATGATTTGCGAAGGGTGAACCTCCGTCGTGTTGCTGCTGTGTACTTATG 1380
 QY 461 GlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeu 480
 DB 1381 GGAAGGGGATTTGAATTAACCTGAGAGGTGCAATGTTTCAACAAGAACTGATTTG 1440
 QY 481 AsnMetLeuAlaLysHisTYrAlaValAlaIleGlyCysGluIleValaPhePheValPro 500
 DB 1441 AATATGCTCCCTTAACCTTACGCCCTGTCATTTGAGAGTGAAGTCCATTTTTCGATACA 1500
 QY 501 AspArgGluGluAspPheAlaSerTYrThrgLysPheLeuArgTYrLeuSerSerLysAsp 520
 DB 1501 GACAGGGAAGAAATTTTGGCTTTCACATGAAATTTCTCCGGATCTTGGCTCAAAAGAT 1560
 QY 521 ArgAlaGlyValAlaLysLeuAspArgGlyThrThrLeuPheLeuValaProProSerAsp 540
 DB 1561 CGGCGGGGTGTTCCAAATTAAGTATGATGATCAAACTTATCTTGGTGTGCTTCATCAGAT 1620
 QY 541 PheLeuThrAspValaLeuGlnValaThrArgGlnGluArgLysTYrGlyValaValaLys 560
 DB 1621 TTCTTAACGATGATCTCAAGTGAACCTCAAGAACGCTTATATGAGTTTGTTCATAG 1680
 QY 561 LeuProProProAlaValProValThrAlaSerTYrArgGlnGluSerGlnSerAsnPro 580

```

Db      1681 TTAACCCCGGCGAGCGTCTGTTACAGATCATATACAGACAAATCTCAGTCCATCTT 1740
Qy      581 LeuHisTyrMetAspGlnAlaIatgAspSerProAlaAsnAlaSerHisSerLeuTyrPro 600
Db      1741 CTGCATTAATATGATGATCAGCCCGGAGTTCACCTGCCAAATCTGATCAGATTATATCTCT 1800
Qy      601 ProArgGluAsnTyrTrileArgGlyAlaProGluHisLeuThrAlaIaSerLysProSer 620
Db      1801 CCGAGGAAATTAATCAATTAGGGGTGACCAAGAACATTGACAGCTCTTCAAAAACATCT 1860
Qy      621 ValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640
Db      1861 GTTAGGAGACCTCTCAGATTAATTAATATGAGCCCTCAAGCTGGGTTAGTTAACT 1920
Qy      641 ProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro 660
Db      1921 CCGAGACTTATAGCCACTCTGGCATCTATCTCCCTGCAAACTTCTCAACCTGCTGCCCT 1980
Qy      661 GluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680
Db      1981 GAGAGTCACCAACCTATGTCAGACCTTCAACAGTTGTTCCACAGCACATCAGTCCAAAT 2040
Qy      681 GlyLeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHis 700
Db      2041 GGAAGCTTACATGAGAGAACGACCTCTCAAGCTTGAAAAAGAGTCCACAAACAGTTCAAT 2100
Qy      701 AspAlaSerAsnGlnSerPheGlnGlnTyrGlyValAsnGlnTyrThrProAlaGlyGlnLeu 720
Db      2101 GATGCTCAATATGATCATTTCCAAATATCGGAATCAGATCAGTCCAGCTGGGCAACTA 2160
Qy      721 ProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMet 740
Db      2161 CCTCCTCTCTCTCTGCGTTACCTCTCAAGCTTCAACCAACCCCACTACACTAGTGAAGT 2220
Qy      741 ValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeu 760
Db      2221 GTTCAGAGGCAACATGCAATACAGAACCATCTGTTAACTGCTGCTGCTCCGTTA 2280
Qy      761 ProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProVal 780
Db      2281 CCAAAATATGCTCATATATATATATTCATGATACACTCAGGGTTCTGCAATCATCTCTGTT 2340
Qy      781 SerGlnProMetValGlnGlnTyrGlnProGlnAlaSerMetProAsnGlnAsnTyrGly 800
Db      2341 TCTCAGCCCATGTCAGCAATACCAACCAAGAGCGTCCATGCAAACTATGCT 2400
Qy      801 ProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn 820
Db      2401 CCAATTCAGATTATACAGAGCTAATTTTTCATGCGGTAAACAACAATCAGGCACAGAAC 2460
Qy      821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGln 840
Db      2461 TTTAAACCTTCCCAATTTTCAAGCTGCAATGCAACCAAGCATATAGGAAATTTTAGAG 2520
Qy      841 ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyTyrThr 860
Db      2521 CCACAAACCAACAGACTTACGATTTGAGCGCTATATCTCTGGGATGCTCAGGGTACAAAC 2580
Qy      861 AspGlyGlyValAspLysAsnGlnAsnArgTyrGlnSerThrLeuGlnPheAlaAsnLeu 880
Db      2581 GATGGGAGAGTGCATATGATCAGATACAGATACAGTCAACATCAATTTTGACAGAACTT 2640
Qy      881 LeuLeuGlnIleGlnGlnTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 900
Db      2641 CTCTCTCAATACAGAGAAACAGCAGCAACAGTCTTCAAGTACTCCGGCTGACAGGGG 2700
Qy      901 Pro 901
Db      2701 CCT 2703

```

RESULT 2
BT003029

```

LOCUS      BT003029                2706 bp    mRNA       linear    PLN 16-JAN-2003
DEFINITION Arabidopsis thaliana Atcg43410/T1024.15 gene, complete cds.
ACCESSION  BT003029
VERSION    BT003029.1
KEYWORDS   FLI CDNA.
SOURCE     Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE  1 (bases 1 to 2706)
            Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Natsuka,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K.,
            Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
            Ecker,J.R.
            Arabidopsis ORF clones
            Unpublished
            2 (bases 1 to 2706)
            Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Natsuka,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K.,
            Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
            Ecker,J.R.
            Direct Submission
            Submitted (16-JAN-2003) Salk Institute Genomic Analysis Laboratory
            (SIGNAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT    RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN
            Arabidopsis Full-length cDNA') : Seki,M., Natsuka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.
            The Salk, Stanford, PGSC (SSP) Consortium members constructed and
            sequenced the pUNI (ORF) clones using the RAPL cDNAs: Cheuk,R.,
            Chen,H., Kim,C.J., Shinn,P., Bowser,L., Chan,M.M., Chang,C.M.,
            Dale,J.M., Hsuan,V.W., Jones,T., Karlin-Neumann,G., Lam,B.,
            Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J.,
            Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
            Yamada,K., Yu,G., Yuan,S., Davis,R.W., Theologis,A., and
            Ecker,J.R.
            Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES   Location/Qualifiers
            source          1..2706
                                /organism="Arabidopsis thaliana"
                                /mol_type="mRNA"
                                /db_xref="taxon:3702"
                                /chromosome="2"
                                /clone="U21966"
                                /ecotype="Columbia"
                                /note="This clone is in pUNI 51"
            CDS              1..2706
                                /note="unknown protein"
                                /codon_start=1
                                /product="Atcg43410/T1024.15"
                                /protein_id="At2G3594.1"
                                /db_xref="GI:27764946"
                                /translation="MALSMKPRADDSGFQSNIMVGLTBTETTSIDTLBFGYDI
                                DRTIVSSRGEAFIYRHEVAVAKELQANLNGSQIKIAYAPAPCSSLWGI
                                GPNVSKDLSEFEFSFKGIEDFRFLREKRTAFIDYENDMDLQASNPKWGSFLR
                                VDFLSQAPKKEOMAGSYDNRNNGNNHNRKPYHSEYEDKGVDPKVLMIIGPPYATQ
                                CNDQOILHNMILFGEIERIVSYSPRNALVBFRAEERPKCEBLOQLRFNNPIKI

```

MYNDELPEODDTFSYSGMKRSRTDMFNNDPSFVSSPHSTGI PGSMRPRGRNERSY
 NGAEYNDVVGKPMWRBPSANGTGLISPGSGILBPPAGTRPRMKSNDSEGDIP
 AOIYREKRRPRDSDVDGFTPMGVDSFRGGSVAAPRIGPPSDBIWGMIAKGT
 PVCCARCPMKGIETLPEVNCSTADTLMKHYAVALGEIYEVFREDEPAST
 YTEFLRYSKDRAGVAKLDGDTLFLVPSDEFLVDLQVTRBERLAKIPPAV
 PVTSYQESQSNLHAYMDQADRS PANASHLYPRENYRGAPEHLTAASKSVSE
 LRITNNAPQAGVSLTEPELLATLASIIPATSOAPAESHOMGSPSTVSTANOSGL
 YNGEAPSOAMKRGPOTVHASNOSFOOYGQYTPAGQLPPSPRYPASNPYTSGM
 VHGMOYOSOSVNMPLSP.LPMHNHYSMTGSSNHPISQMVQOYOPBASPMON
 YGPIRSYQANFHVTTTNOANLPSQFQAMQPPAKANLEFONQALRLOPMISDGG
 QGTTDSVDKNQRIQSTLQFPAANLLQIQKQKQSSGTPAGQGP

ORIGIN

Alignment Scores:

Pred. No.:	3.64e-183	Length:	2706
Score:	4786.00	Matches:	899
Percent Similarity:	99.89%	Conservative:	1
Best Local Similarity:	99.78%	Mismatches:	0
Query Match:	99.69%	Indels:	0
DB:	8	Gaps:	0

US-09-920-705-3 (1-901) x BT003029 (1-2706)

QY 1 MetAlaLeuSerMetLeuProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
 Db 1 ATGGCGTTATCTATGAGCCATTCAGAGCCGATGATCCGGTTTCCAGTCAACATCTT 60
 QY 21 TrpValGlySerLeuThrProGluThrGlnSerAspLeuThrGluLeuPheGlyArg 40
 Db 61 TGGGTCGGTGACCTTACGCCGAGACGACAGAGTCAAGTCTGACCAAGTGTGGAGACA 120
 QY 41 TyrGlyAspIleAspArgIleThrValIlyrSerSerArgGlyPheAlaPheIleTyrTyr 60
 Db 121 TACGGCGCATATGTATGAAATCACGGTATCTTCAAGAGCGCTTCGTTTAAATATCTAC 180
 QY 61 ArgHisValGluGluValAlaAlaIlyrGluValAlaGlnGlyValAlaAsnLeuAngly 80
 Db 181 AGCAGTGTGAGAGAGAGAGTGGCAGCAAGAGGCTCTTCAAGAGCAAAATTTGATGCA 240
 QY 81 SerGlnIleIlyrIleGluTyrAlaArgProAlaIlyrProCylSerSerLeuThrValGly 100
 Db 241 AGTCAATTAAGTCAATTCAGACGACGCGCAAACTTGTAAAGTCTTAAGGTGGGT 300
 QY 101 GlyIlyleGlyProAsnValSerIlyrAspAspLeuGluGluLeuPheSerIlyrAspGly 120
 Db 301 GGAATGGCGCTTAATCTCCAGAGTGAAGTCTGAGAGAGAGTTCAGCAAGTTGGGAAA 360
 QY 121 IleGluAspPheArgPheLeuArgGluArgIlyrThrAlaPheIleAspTyrTyrGluMet 140
 Db 361 ATCGAGGATTTTNGGTTCTTCAGAGACGCAAGACGCTTCAATGATTAATATAGATG 420
 QY 141 AspAspAlaLeuGlnAlaIlyrSerMetAsnGlyIlyrProMetGlyGlySerPheLeuArg 160
 Db 421 GATGATGCTTTACAGGCTAAGACATGATGAAAGCCATAGGAGTTCCTTTTGGCT 480
 QY 161 ValAspPheLeuArgSerGlnAlaProIlyrIlyrGluGlnIleThrAlaGlySerTyrAspAsn 180
 Db 481 GTTGAATTTCTCCGGTCACAGCGCCCAAAAAAGAACATGGCGCTCTTACCATTAAC 540
 QY 181 ArgAsnGlyAsnMetAsnHisIlyrSerProGlnTyrProHisSerTyrGluAspPheIlyrGly 200
 Db 541 AGAATGGCAATATGATATTAACCGCATATCTCTCACTCATATGAAGACTTTAAAGCA 600
 QY 201 AspValGlnProSerIlyrValLeuThrIleGlyPheProProThrAlaThrGlnCysAsn 220
 Db 601 GATGTCGAGCCCAAGTAAAGGTTCTGTGATTTGGTTCCTCCCTCACTGCTAACATGCAAT 660
 QY 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValIlyrSer 240
 Db 661 GATGACCAATTTCTCAACAATGCAATGATCTCTTGTGTGAGATCGAAGAGGTAAAGT 720
 QY 241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260

Db 721 TACCATCAAGAAATTTTGCACCTTGAGATTAGAGCGCGAGAGAACTCCCAATGC 780
 QY 261 IlyrGluGlyLeuGlnGlyYArgLeuPheAsnAsnProArgIleIlyrIleMetTyrSerAsn 280
 Db 781 AAGGAAGGCTTACAGGAGAGGTTATTCATTAATCTTAAGATCAAAATTAATGACTCAAC 840
 QY 281 AspGluLeuProProGluGlnIleAspAspThrSerPheTyrSerGlyMetIlyrAspSerArg 300
 Db 841 GATGATGCTCTCCAGAGCAACATCTATGTTTCTCTGATTAACCGTCAAG 900
 QY 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
 Db 901 ACGATATGTTCAATTAATGATCTTCAATTTGATCTTCCCATCTACAGCAATTCCT 960
 QY 321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyValAlaIlyrCysAsn 340
 Db 961 GGGTCTAAGAGGCGCCCTCAGAGGTACAAATAGGGCTCATTAATGGTCAAGATACAT 1020
 QY 341 AspValValGlyIlyrGluProAsnThrArgArgProSerAlaAsnGlyTyrGlyIleLeu 360
 Db 1021 GACGTTGTTGGTAAAGAGCCAACTGAGAGGCGCATCTGCAAAATGGAATGAACTACTC 1080
 QY 361 ProSerProThrGlyProGlyIleLeuProSerProAlaGlnIlyrThrArgArgProMet 380
 Db 1081 CCATCTCAACAGACCTGGAATCTCTCCATCTCTGCAAGGTACAGAGCGCTATG 1140
 QY 381 ArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerIlyr 400
 Db 1141 AGGTCAAAACCCGATCTTGGGAAGATATGATCTCTGCTGATGGTCAAGAAAGTAA 1200
 QY 401 ArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArgSer 420
 Db 1201 CGAACCAAGAGATGATCGATGAGTGAAGTTCCTCAATGGGTGTCGATGAGAGGTCA 1260
 QY 421 PheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHisIle 440
 Db 1261 TTTGGTGAAGTTCAAGTTGCTGCTACACATACGATGCGCCCTGATTTCTGATACATA 1320
 QY 441 TrpArgGlyMetIleAlaIlyrGlyIlyrThrProValCysCysAlaArgCysValProMet 460
 Db 1321 TGGAGAGAAAGATTTGCGAAGGGTGAACCTCCGTGTTGTCTCTGTTGTATGACTATG 1380
 QY 461 GlyIlyrGlyIleGluTyrIlyrLeuProGluValValAsnCysSerAlaArgThrAspLeu 480
 Db 1381 GGAAGGAGATTAACCTTAACCTGAGAGTCTCAATTTCTTCAAGCAACCTGATTTG 1440
 QY 481 AsnMetLeuAlaIlyrHisTyrAlaValAlaIleGlyCysGluIleValPhePheValPro 500
 Db 1441 AATATGCTCGCTAAACATTAACGCGTTCATTTGATGATGATCGTTTTTTCGTACCA 1500
 QY 501 AspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerIlyrAsp 520
 Db 1501 GACAGGAGAAAGATTTTGGCTTACACGAAATTTCTCCGATACCTTGAACCAAAAGAT 1560
 QY 521 ArgAlaGlyValAlaIlyrLeuAspAspGlyThrThrLeuPheLeuValProProSerAsp 540
 Db 1561 CGGGCGGGTGTGCCAAATTAAGATGATGATCAACTTATCTTGGTGTCTCATATGAT 1620
 QY 541 PheLeuThrAspValLeuGlnValThrArgGlnIlyrArgLeuTyrGlyValValLeuIlyr 560
 Db 1621 TTTCTTAATGATATGATCTCCAGAGTGAACCGTCAAGAACGCTATATAGTGTGTTCAG 1680
 QY 561 LeuProProProAlaValProValThrAlaSerTyrArgGlnIlyrSerIlyrAsnPro 580
 Db 1681 TTACCCCGGCAAGCGCTTCTGTACAGATCAATCAAGCAAGAACTCAAGTCAATCT 1740
 QY 581 LeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrPro 600
 Db 1741 CTGCAATTAATGATCAAGCCCGGATTCACCTGCCAAAGCTAGTCAAGTTTAAATCT 1800
 QY 601 ProArgGluAsnTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSerIlyrProSer 620
 Db 1801 CTTAGGAAAAATTAATTAAGGGGTGCAACCAAGCAATTTGACAGCTGCTTCAAAACATCT 1860

QY 621 ValSerGluProLeuValGlyLeuProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640
 Db 1861 GTTAGGAGACCTCTCGAAATACCTTAATATGACGCCCTCAAGCTGGGTGATTAACT 1920
 QY 641 ProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro 660
 Db 1921 CCGGAGCTTTAGCCACTGCGATCTATTCCTCCGCAACTTCTCACTGCTGCCCCCT 1980
 QY 661 GluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680
 Db 1981 GAGAGCACCAACCTATGTCAGGACCTCAACAGTTGTTCCACAGACATCAGTCCAAAT 2040
 QY 681 GlyLeuTyrAsnGlyValAlaProSerGlnAlaAlaThrValArgGlyProGlnThrValHis 700
 Db 2041 GGAAGTACAAATGGAAGACAGCGTCTCAAGCTTGGAAGAGAGTCCACAAACAGTTCAT 2100
 QY 701 AspAlaSerAsnGlnSerPheGlnGlnTyrGlyValAsnGlnTyrThrProAlaGlyGlnLeu 720
 Db 2101 GATGCTCAATCAGTCACTTCCAAATACGGAATCAGTACCTCAGCTGGGCAACTA 2160
 QY 721 ProProProSerSerArgTyrProProAlaSerAsnAsnProAsnTyrThrSerGlyMet 740
 Db 2161 CTTCCCTCTCTCGCTGCGGTACCTCCAGCTTCAACACCCCACTACACTAGTGAATG 2220
 QY 741 ValHisGlyAsnMetGlnTyrGlnSerGlnSerValAsnMetProGlnLeuSerProLeu 760
 Db 2221 GTTCCATGAGCAACATGCAATACCAAGACCAATCTGTTAACATGCTCAGTCTCGGTTA 2280
 QY 761 ProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProVal 780
 Db 2281 CCAAAATATGCTCATATATATATTTTCCATGATACCTCAGGGTTCTCAATCATCTCTGT 2340
 QY 781 SerGlnProMetValGlnGlnTyrGlnProGlnAlaSerMetProAsnGlnAsnTyrGly 800
 Db 2341 TCTCAGCCCATGGTCCAGCAATTCACCAACGAAAGCTCCATGCAACCAAACTATGAT 2400
 QY 801 ProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn 820
 Db 2401 CCAATTTCCAAATGATACGAAAGCTAATTTTCATGGGCTAACAACAAATCAGGCAAGAAC 2460
 QY 821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspValAsnLeuGln 840
 Db 2461 TTAAACCTTTCCAAATTTCAAGCTGCGCATGCAACCAACGACATAGGCAATTTTAAAG 2520
 QY 841 ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThr 860
 Db 2521 CCACAAAAACCAAGCACTACGATTCAGCTATGATCTCTGGGATGCTCAGGGTACAAACA 2580
 QY 861 AspGlyValValAspValAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaLeu 880
 Db 2581 GATGGGAGAGTGCATAGATCAGATACCACTCAACTCAATTTGACAGCAAACTT 2640
 QY 881 LeuLeuGlnIleGlnGlnTyrGlnGlnGlnGlnGlnSerSerGlyThrProAlaGlyGlnGly 900
 Db 2641 CTTCTTCAATACAGCAAAACGCGACAGCAACAGTCTTCAAGTACTCCGCTGACAGGGG 2700
 QY 901 Pro 901
 Db 2701 CCT 2703

RESULT 3
 LOCUS AY094463 3498 bp mRNA linear PLN 05-MAY-2002
 DEFINITION Arabidopsis thaliana At2g43410/71024.15 mRNA, complete cds.
 ACCESSION AY094463
 VERSION AY094463.1 GI:20453184
 KEYWORDS FLI CDNA
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Bukayeva, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE
 AUTHORS
 1 (bases 1 to 3498)
 Shin, P., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Ban, J.,
 Bower, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D.,
 Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G.,
 Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M.,
 Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T.,
 Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C.,
 Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinzaki, K., Davis, R.W.,
 Theologis, A., and Ecker, J.R.
 Arabidopsis cDNA clones
 Unpublished
 2 (bases 1 to 3498)
 Shin, P., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Ban, J.,
 Bower, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D.,
 Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G.,
 Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M.,
 Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T.,
 Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C.,
 Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinzaki, K., Davis, R.W.,
 Theologis, A., and Ecker, J.R.
 Direct Submission
 Submitted (08-APR-2002) Salk Institute Genomic Analysis Laboratory
 (Signal), Plant Biology Laboratory, The Salk Institute for
 Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
 USA
 COMMENT
 JOURNAL
 TITLE
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RFLP cDNAs: Shin, P., Chen, H.,
 Arabidopsis Full-length cDNA: Seki, M., Narusaka, M., Ishida, J.,
 Hayashizaki, Y., and Shinzaki, K.
 The Salk, Stanford, PGSC (SSP) Consortium members carried out the
 sequencing and annotation of the RFLP cDNAs: Shin, P., Chen, H.,
 Cheuk, R., Kim, C.J., Meyers, M.C., Ban, J., Bower, L., Chang, E.,
 Dale, J.M., Goldsmith, A.D., Jones, T., Karlin-Neumann, G., Lam, B.,
 Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J.,
 Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C.,
 Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W., Theologis, A.,
 and Ecker, J.R.
 Shin, P., (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to
 this work. Shinzaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
 contributed equally to this work as PIs.
 FEATURES
 source
 location/Qualifiers
 1..3498
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="2"
 /clone="RAF109-47-P10(R121666)"
 /ecotype="Columbia"
 255..2960
 /note="unknown protein"
 /codon_start=1
 /product="At2g43410/71024.15"
 /protein_id="AA119833.1"
 /db_xref="GI:20453185"
 /translation="MALSKPEFRADDSGFQSNLNVGSLTPETTESDITELFGRYGDI
 DRITVSSRGFAFYIRHVEBAVAKKALQNGSOIKIEYARAPARCKLAWGDI
 GNVSKDLSEEFSEKLEIDREFLRKERTAFIDYEMDDALQAKSGSFLR
 VDFLRSAQPKKQMGSYDNRNGNNHPOYHSHVEDKDPQSPKVLWIGPPRATO
 CNDDEIILNANILFGEIERVYSYPERNATLVEFRSEARCKEQLQRLFPNNPIKI
 MTSNDELPEEDDITFYSKMSKRTDMRNDSFVSPHSTGIPEMSPFLRTNBSY
 NGAENVDVAGEKPNMRKANGCTGILPPTGEGILPSPAQGRPRMRSPDSWEGSDP
 AOLVSEKRTKRGSDVDGFTTNGVDERSFSGSVARPIRGPDSDHLMKGMIAAGCT
 PYCARCVPMGKGIETKLPEVYNGSARFDLNNLAGHVAALGCEIVFVDEEDFPAS
 YTEFLRYLSSDRAGAVATLDGTTTLFLVPSPLFDVLOVTRQELHLYGVKLPPPAV
 PRTASVROSGSNPLHYMDQARDSPANNAHSHTYPRNVTIGAPHLTAASPSVSP
 LATIPNNAAPQAGVSLTPELALATLASLTPATSGPAPAPESHQPSGSTVSTHQSNGL
 VNGEAPSQAMKRGQPTVHADSNQSFQYQNGTPTGQLPPEPSRYPASNNPNYSNM
 VHGNNQYOSGVNMPQLSLPLPMPIRNNYSMTYQGSNNHVSQPMVQYQVPEASMPNM
 YGPISTYQOANHGVTNQAOLNPSQFOQAAQMPADKANLEPQANLRLQPMISGDG
 GCTTGCVKDKNRVYSTIQLPANNLLQLQKRRQGSQSGRPAGQP"

misc_difference 2917
/note="compared to genome"
/replace="a"

ORIGIN

Alignment Scores:

Pred. No.:	4,78e-183	Length:	3498
Score:	4786.00	Matches:	899
Percent Similarity:	99.89%	Conservative:	1
Best Local Similarity:	99.78%	Mismatches:	1
Query Match:	99.69%	Indels:	0
DB:	8	Gaps:	0

US-09-920-705-3 (1-901) x AY094463 (1-3498)

QY 1 MetAlaLeuSerMetLysProPheAArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20

DB 255 ATGGCGTTATCTATGAAAGCCATTCAGAGCCGATGATTCGCTTCCAGTCAAACTTCTT 314

QY 21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40

DB 315 TGGGTGGGTAGCCCTAAGCCGAGAGACAGAGTCAGATGTGACCAAGTTGTTGAGAG 374

QY 41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60

DB 375 TACGGCGATATGATGAAATCACGGTGATTTCTTCAAGAGCTTTCGCTTATATATCTAC 434

QY 61 ArgHisValGluGluAlaValAlaAlaIleValGluAlaLeuGlnGlyValAsnLeuAsnGly 80

DB 435 AGACATGTGGAGAGACAGTCGACGCCAAGAGGCTTTCAGAGCAAAATTTGAATGGA 494

QY 81 SerGlnIleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTrpValGly 100

DB 495 AGTCAATTTAAGATTCGAATTCAGCAGACCGGCAAACTTTGTAAGAGTCTATGGGTGG 554

QY 101 GlyIleGlyProAsnValSerLysAspAspLeuGluGluPheSerLysPheGlyLys 120

DB 555 GGAATGGCCCTTATATCTCCCAAGATGACCTGGAGAGAGATTCAGCAAGTTGGGAAA 614

QY 121 IleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMet 140

DB 615 ATCGAAGATTTTAAAGTTCTCGAGAGAACCAAGACAGCTTCAATGATATATATGAGAT 674

QY 141 AspAspAlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArg 160

DB 675 GATGAGCTTTTACAGGCTTAAGCATGGAATGGAAGCTTAAGGTGATGCTTTTGGCT 734

QY 161 ValAspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsn 180

DB 735 GTTGAATTTCTCGGTCAAGGCCCAAAAAGAAACAATGGGCTGATTCATGATTAAC 794

QY 181 ArgAsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGly 200

DB 795 AGAAATGCAATATGATCATTAACCGGATTCCTCACTCATATGAGAGCTTTAAAGGA 854

QY 201 AspValGlnProSerLysValLeuTyrIleGlyPheProProThrAlaThrGlnCysAsn 220

DB 855 GATGTCACCAAGTAAAGGTCTGTGAGATTTGGGTTCCCTCACTCACTCAACAATGCAAT 914

QY 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240

DB 915 GATGAGCAAAATTTCTGACAAATGCGATGATCTTTGGTGAAGTGAAGGGGTAAAGT 974

QY 241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260

DB 975 TACCATCATCAAGAAATTTTGCACTGTGGAGTTTAAAGAGCGGAGAAAGCTGCCAAAGC 1034

QY 261 LysGluGlyLeuGluGlyArgGluPheAsnAsnProAlaGlyIleValIleMetTyrSerAsn 280

DB 1035 AAGGAAGGCTTACAGAGGAGGTATTTCAATATCTTAAGATCAAAATTAATGACTCAAC 1094

QY 281 AspGluLeuProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300

DB 1095 GATGAGTTCCTCTGAGCAGACATCACTAATTTTACTGTGTATGAAACGCTCAAG 1154

QY 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320

DB 1155 ACAGATATGTTCAATATATGATCTTCAATTTGATCTTCTCCATCTTCACTGAAATTCCT 1214

QY 321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyValGlyTyrAsn 340

DB 1215 GGGTCTATGAGGCCCTCCAGAGGTCAAAATGAGCTTATATTAATGTCAGAAATACAT 1274

QY 341 AspValIleGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeu 360

DB 1275 GACGTTGTTGGTAAAGACCAAACTGAGAGGCCATCTGCAATAGGAATCGAAATCTC 1334

QY 361 ProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMet 380

DB 1335 CCATCTCCAAACAGACACTGGAATCTCCATCTCTGCAAGAGTACAGAGCGCCATATG 1394

QY 381 ArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGlnLeuValArgGluSerLys 400

DB 1395 AGGTCAAAACCCGATTTTGGGAAGATATATCTGCTCAGTTGCTCAGAGAAAGTAA 1454

QY 401 ArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArgSer 420

DB 1455 CGAACCAAGAGAGATGATCACTGACCGTTTACTCCAAATGGGTTCGATGAGAGTCA 1514

QY 421 PheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHisIle 440

DB 1515 TTTGCTCAGAGTTCAGTTGCTGCTAAGCTATCCGTGGCCCTCGATTCTGATCACA 1574

QY 441 TrpArgGlyMetIleAlaLysGlyGlyThrProValCysCysValAspCysValProMet 460

DB 1575 TCGAGAGAAATGATTTCCAGAGGTGAGACTCCGCTGTTGTGCTGTTGTTGTTGTTG 1634

QY 461 GlyLysGlyIleGluThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeu 480

DB 1635 GGAAGGGGATTTGAACCTTAACCTCGAGGTGCTCAATTTTGAAGCAAGACTGATTG 1694

QY 481 AsnMetLeuAlaLysHisTyrAlaValAlaAlaIleGlyCysGluIleValPhePheValPro 500

DB 1695 AATATGCTCGGTAAACATTCAGCCGTTGCCATTTGGAGTGAAGTCGTTTTTTCGATCA 1754

QY 501 AspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAsp 520

DB 1755 GACAGGGAAGAAATTTTGGCTTCACTGAATTTCTCCGGTACCTTCACTCAAAAGAT 1814

QY 521 ArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerAsp 540

DB 1815 CGGCGGGGTGTGCAATTAATGATGATGATCAACTTATTTTGGTGGCTCCATCAGAT 1874

QY 541 PheLeuThrAspValIleGluGlnValThrArgGlnGluArgLeuTyrGlyValIleLeuLys 560

DB 1875 TTCTTAATGATGATCTCAAGTGAACCGTCAAGAACGGCTAATATGTTGTTCTCAAG 1934

QY 561 LeuProProProAlaValProValThrAlaSerTyrArgGlnGluSerGlnSerAsnPro 580

DB 1935 TTACCCCGCAGCGCTTCTGTTAAGCATCATTAAGACAAABAATCTCACTCAATCCT 1994

QY 581 LeuHisTyrMetLeuArgGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrPro 600

DB 1995 CTGATTAATATGATCAAGCCCGGATTCACCTGCCAAATGCTAAGTCAAGTTATATCT 2054

QY 601 ProArgGluAsnTyrIleArgGlyValIleProGluHisLeuThrAlaAlaSerLysProSer 620

DB 2055 CTRAGGAATTAATTAATGAGGTGACCAAGAACATTTGACGCTCTTCAAAACATCT 2114

QY 621 ValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640

DB 2115 GTTAGAGAGGCTCTCAGAAATACCTAATATGACAGGCTCAAGCTGGGTATGTTAACT 2174

QY 641 ProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro 660

DB 2175 CCGAGACTTTTACCACTCGGATCTATTTCTCCGCAAACTTCTCAACCTGCTGCCCT 2234

```

QY 661 GUserHiGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680
DB 2235 GAGAGTCACCAACCTATGTCAGACCTTCACAGCTGTTCCACAGCACATCAGTCAAT 2294
QY 681 GLeuTYrAsnGlyValAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHis 700
DB 2295 GGAAGTCTACATGAGAGACAGCCGTCTCAAGCTTGAGAAAAGAGTCCACAAACAGTTCT 2354
QY 701 AAspAlaSerAsnGlnSerPheGlnGlnTYrGlyAsnGlnTYrThrProAlaGlyGlnLeu 720
DB 2355 GATGCGTCAATCAGTCAATCCACACATACGAGAAATCAGTACACTCCAGCTGGGCAACTA 2414
QY 721 ProProProProSerTrpTYrProProAlaSerAsnAsnProAsnTYrThrSerGlyMet 740
DB 2415 CCGCTCCCTCCCTTGGGTTACCTCCAGCTTCAAAACCCCAACTACACTATGTGAAAG 2474
QY 741 ValHisGlyAsnMetGlnTYrGlnSerGlnSerValAsnMetProGlnLeuSerProLeu 760
DB 2475 GTCCATGCGCAATGCAATACAGAGCCCAATCTGTTAACTGCTTCAAGCTGTCTCCGTTA 2534
QY 761 ProAsnMetProHisAsnAsnTYrSerMetTYrThrGlnGlySerSerAsnHisProVal 780
DB 2535 CCAATATGCTCATATATATATATTCATGTAACACTCAGGGTTCGTAATCATCTCTGTT 2594
QY 781 SerGlnProMetValGlnGlnTYrGlnProGlnAlaSerMetProAsnGlnAsnTYrGly 800
DB 2595 TCTCAGCCCATGCTCCAGCAATTCACCAACGAGACGCTCCATGCAACCAAACTATGAT 2654
QY 801 ProIleProSerTYrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsn 820
DB 2655 CCAATTCAGATTATACAGCAAGCTAATTTTCATGGGTTACACAAATCAGGACAGAAC 2714
QY 821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGln 840
DB 2715 TTAACACCTTCCCAATTTTCAAGCTGCCATGCAACCAACAGATTAAGGCAATTTAGAG 2774
QY 841 ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyTYrThr 860
DB 2775 CCACAAACCCAGCACTAGCATTCAGCCCTATGATCTCTGGGATGCTCAGGGTACAAACA 2834
QY 861 AAspGlyValAlaAspLysAsnGlnArgTYrGlnSerThrLeuGlnPheAlaAlaAsnLeu 880
DB 2835 GATGGGAGAGTCCATTAAGATTCAGAGATTCACACTCACTCAATTTGCAAGCAAACTT 2894
QY 881 LeuLeuGlnIleGlnGlnLeuGlnGlnGlnGlnSerSerGlyThrProAlaGlyGlnGly 900
DB 2895 CTTCTCCAGTACAGCAAGAAACGCGACAGCAACATCTTCAAGGTACTCCGGCTGCAAGGG 2954
QY 901 Pro 901
DB 2955 CCT 2957

RESULT 4
LOCUS CQ804588 3171 bp DNA linear PAT 10-MAY-2004
DEFINITION Sequence 999 from Patent WO2004035798.
ACCESSION CQ804588
VERSION CQ804588.1 GI:47110928
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE
1 Inze, D., de Veylder, L. and Vliethe, K.
AUTHORS Identification of novel e2f target genes and use thereof
JOURNAL Patent: WO 2004035798-A 999 29-APR-2004;
CROPDESIGN N.V. (BB)
FEATURES
location/Qualifiers
1..3171
/organism="Arabidopsis thaliana"

```

```

ORIGIN .. /mol_type="unassigned DNA"
/db_xref="taxon:3702"

Alignment Scores:
Pred. No.: 6,6e-178 Length: 3171
Score: 4656,50 Matches: 892
Percent Similarity: 85,04% Conservative: 6
Best Local Similarity: 84,47% Mismatches: 3
Query Match: 96,99% Indels: 155
DB: Gaps: 3

US-09-920-705-3 (1-901) x CQ804588 (1-3171)

QY 1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
DB 1 ATGGCGTATCTATATAGCATTCAGAGCCGATGATTCGGTTCACAGCAAACTATCT 60
QY 21 TrpValGlySerLeuThrProGlnThrTrpGlnSerAspLeuThrGluLeuPheGlyArg 40
DB 61 TGGGTGGTAGCTTAACCGCGAGACGACAGATCAATCTGACGAGTTGTTGAGACA 120
QY 41 TYrGlyAspIleAspArgIleThrValTYrSerSerArgGlyPheAlaPheIleTYrTYr 60
DB 121 TAGGGGATATGATGATGATCAAGGTATTTCTTACAGAGGCTTGGCGTTATATATATAC 180
QY 61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuGlnGly 80
DB 181 AGACATGTGAGAGGAGAGAGTCCAGACCAAGAGGCTCTTCAAGAGGCAAAATTTGATGCA 240
QY 81 Ser----- 81
DB 241 AGCAATGAGCTCTGCAAAAGAAATACATCAACACATGTTTACTCTCGTTTACCAAGAA 300
QY 81 ----- 81
DB 301 GGTTAATCCATTTGTCGTTATCTGCAAACTGTTATCCGCTTGAAGTCTGATTTCA 360
QY 81 ----- 81
DB 361 GGTTCCTCAGATGTTAATCTCAACAGCTTGAATGATGATGTTGGTACTCCATC 420
QY 82 -----GlnIleLysIleGluTYr----- 87
DB 421 GAAATTCGATGATGAGAAATTTTGGAGTTTAAAGTTCAAGTTTCTGCAAAATCAAA 480
QY 87 ----- 87
DB 481 CCGCGTGAAGAAATTTGTCTATGGGGTACGTAGATATCAATATACATCTGCGAACA 540
QY 87 ----- 87
DB 541 GAAATTCGCAACCGTATCGTCTTAATCTAGATGCTGAGACATATCCATATCAG 600
QY 87 ----- 87
DB 601 TGTGAGGAGATTGGGCTTGTGTTGGCTCGATGCTTCAGAGAACAAATTAAGTGGTTG 660
QY 87 ----- 87
DB 661 TTAAGCGGCTAATCTACCATCAGAAACGTTTATTCGAGAAACATGTTGTTCTTC 720
QY 88 ---AlaArgPro---AlaLysProCysLysSerLeuTrpValGlyGlyTYrProAsn 105
DB 721 AATTCACATCCCTCAGCAAAACCTGTAAAGATCTATAGGTGGGTGGAATCGGCTTAAT 780
QY 106 ValSerLysAspAspLeuGlnGlnGlnLysPheSerLysPheGlyLysIleGlnAspPheArg 125
DB 781 GTCTCCAGAGAGACCTTGAGGAGAGAGTTTCACCAAGTTGGGAAATTCAGAGATTTTAG 840
QY 126 PheLeuArgGluArgLysThrAlaPheIleAspTYrTYrGlyLysMetAspAspAlaLeuGln 145
DB 841 TTTCTCAGGAACGCAAGACAGCTTCAATGATTTATGAGATGATGATGATGATGATGATGAT 900

```

QY 146 AlalysSerMetAsnGlyLysProMetGlyGlySerPheLeuValAspPheLeuArg 165
 DB 901 GCTAAGACATGATGAAAGCCTTAGGGGTAGCTTTTGGCCGTGTGATTTTCTCCGG 960
 QY 166 SerGlnAlaProLysLysGluGlnTrpAlaGlySerTyraAspAsnArgAsnMet 185
 DB 961 TCACAGAGCGCAAAAAAGAACATGGGCTGGCTTACGATACCAAAATGGCAATATG 1020
 QY 186 AsnHisLysProGlnTrpProHisSerTyraGluAspPheLysGlyAspValGlnProSer 205
 DB 1021 AATATATAACCGCAGATATCCCTCACTCATATGAAACATTTAAAGAGATGTCCAGCCAACT 1080
 QY 206 LysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeu 225
 DB 1081 AAGGTTCTGTGGATTTGGGTTCCCTCCATACCTCAACATGATGACAAATTTCTG 1140
 QY 226 HisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyraProSerArgAsn 245
 DB 1141 CACATCGCATGATACCTCTTGTGGATGAGATGAGAGGGGTAAAGTTAACCATCAAGGAAT 1200
 QY 246 PheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGlyLysGluGln 265
 DB 1201 TTTGCACTTGTGAGATTAGAGGCGGAGAGAGCTGCAATGCAAGAGAGGCTTACAG 1260
 QY 266 GlyArgLeuPheAsnAsnProArgIleLysIleMetTyraSerAsnAspGluLeuProPro 285
 DB 1261 GGGGGGTTATTCAATATCTTAGATCAAAATTAATGATCAAAACATGATGCTCTCT 1320
 QY 286 GluGluAspAspTrpSerPheTyraSerGlyMetLysArgSerTyraAspMetPheAsn 305
 DB 1321 GAGCAACAGATACGTAGTTTAACTCTGTATGAAAACGGTCAAGAGCAAGATATGTTCAAT 1380
 QY 306 AsnAspProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgPro 325
 DB 1381 AATGATCCTTCAATTGATTTCTCTCTCACTTCACTGGAATTCCTGGGCTTATGAGGCC 1440
 QY 326 LeuArgGlyTyraAsnGluArgSerTyraAsnGlyAlaGluTyraAsnAspValAlaGlyLys 345
 DB 1441 CTCAGAGGTACAAATAGCGCTTCAATTAATGTGCAGAAATACATGACCTGTGTGTAG 1500
 QY 346 GluProAsnTrpArgArgProSerAlaAsnGlyTyraGlyIleLeuProSerProThrGly 365
 DB 1501 GAGCCAACTGAGAGAGGCCATCTGCAAAATGAACTGGAATCTCCCATCTCCAAACAGGA 1560
 QY 366 ProGlyIleLeuProSerProAlaGlnGlyTyraArgArgProMetArgSerAsnProAsp 385
 DB 1561 CCTGGAATCCTCCCATCTCTGCAACAGGACAGGCGCCCTATGAGGTCAAAACCCCAT 1620
 QY 386 SerTrpGluGlyTyraTrpProAlaGlnLeuValArgGlySerLysArgThrArgArgAsp 405
 DB 1621 TCTTGGGAAGATATATCTCTGCTCAGTGTGACAGAAAGTAAACGAACCGAAGAGAT 1680
 QY 406 GlySerValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySer 425
 DB 1681 GGATCAGTGGAGCGTTTACTCCAAAGGCTGTCGATGAGAGGTCAATTGGTGAAGTTCA 1740
 QY 426 ValAlaAlaArgProIleArgGlyProProAspSerAspHisIleTrpArgGlyMetIle 445
 DB 1741 GTTGCTGCTAGACATATCCGTGGCCCTCGATTCTGATCAATATGGAGAGATGATT 1800
 QY 446 AlaLysGlyGlyTyraProValCysCysAlaArgCysValProMetGlyLysGlyIleGlu 465
 DB 1801 GCCAAGGTGGAACTCCGCTGTGTGTGTGCTGTGTGTGATCAATATGGGAAAGGGATTAA 1860
 QY 466 ThrLysLeuProGluValValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLys 485
 DB 1861 ACTAAACGCTGAGAGTGTCAATGTTCTCAGCAAAACATGATTGAATATGCTCGCTAA 1920
 QY 486 HisTyraAlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluGluAsp 505
 DB 1921 CATTAAGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980

QY 506 PheAlaSerTyraThrGluPheLeuArgTyraLeuSerSerTyraAspArgAlaGlyValAla 525
 DB 1981 TTTGCGCTTACACTGAAATTTCTCCGGTACCTTAGCTCAAAAGATCGGCGGCTGTGCC 2040
 QY 526 LysLeuAspAspGlyTyraThrIleuPheLeuValProProSerAspPheLeuThrAspVal 545
 DB 2041 AATTATGATGATGTATCAACTTATTTCTGTGGCTCCATCGATGATTTCTTAATCTATGTA 2100
 QY 546 LeuGlnValThrArgGlnGluArgLeuTyraGlyValValLeuLysLeuProProProAla 565
 DB 2101 CTCCAAGTACCGGCTCAAGAAAGGCTATATGTTGTGTTCTCAACTTACCCCGCAAGC 2160
 QY 566 ValProValThrAlaSerTyraArgGlnGluSerGlnSerAsnProLeuHisTyraMetAsp 585
 DB 2161 GTTCTGTATACGATCATACAGACAAATCTCAGTCAATCTCTCATATATATGAT 2220
 QY 586 GlnAlaArgAspSerProAlaAsnHisSerHisSerLeuTyraProProArgGluAsnTyra 605
 DB 2221 CAAGCCCGGATTCACCTGCCAATCTAGTCACTATATATCTCTTACGGAATAATAC 2280
 QY 606 IleArgGlyAlaProGluHisLeuThrAlaAlaSerLysProSerValSerGluProLeu 625
 DB 2281 ATTAGGGGTGCACAGAACATTTGACAGCTGCTTCAAAACATCTGTTAGGAGGCTCTC 2340
 QY 626 ArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAla 645
 DB 2341 AGAATACCTAATATGACGCGCTCAAGCTGAGTTATATCTCTTACGGAATAATAC 2400
 QY 646 ThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGluPro 665
 DB 2401 ACTCGGATCATATCTCCCTGCAACTTCTCAACCTGTGCGCTGAGAGTCAACAACT 2460
 QY 666 MetSerGlyProSerThrValAlaSerThrAlaHisGlnSerAsnGlyLeuTyraAsnGly 685
 DB 2461 ATGTAGGACCTTCAACAGTGTGTTCACAGCAATCATGATCAATGATGATGATGATGAT 2520
 QY 686 GluAlaProSerGlnAlaTrpLysArgGlyProGlnThrValHisAspAlaSerAsnGln 705
 DB 2521 GAGAGACGCTCTCAAGCTTGGAAAGAGTCCAAACAGTTCAATGAGTCAAAATCAG 2580
 QY 706 SerPheGlnGlnTyraGlyAsnGlnTyraTrpProAlaGlyLeuLeuProProProSer 725
 DB 2581 TCATTTCCAAATACGGAATCAGTACATCTCAGCTGAGGCAATCACTCTCTCTCTG 2640
 QY 726 ArgTyraProProAlaSerSerAsnAsnProAsnTyraThrSerGlyMetValHisGlyAsnMet 745
 DB 2641 CTTTACCTTCCAGCTTCAAAACAAACCCCACTACATGATGGAATGTGTCATGCAACATG 2700
 QY 746 GlnTyrglnSerGlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHis 765
 DB 2701 CAATACCAAGCCCAATCTGTTAATCATGCTCAGCTGTCTCGTTACCAATATATGCTCAT 2760
 QY 766 AsnAsnTyraSerMetTyraThrGlnGlySerSerAsnHisProValSerGlnProMetVal 785
 DB 2761 AATATTAATTCATGATCACTCAGGGTTCGTCAATATCATCTGTCTCAGCCCATGCTC 2820
 QY 786 GlnGlnTyrglnProGluAlaSerMetProAsnGlnAsnTyraGlyProIleProSerTyra 805
 DB 2821 CAGCAATACCAACAGAGAGCTCATGCCAAACCAAACTATATGCTCAATTCAGATTA 2880
 QY 806 GlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGln 825
 DB 2881 CAGCAAGCTAATTTTCAATGCGTATACCAACAAATCAGGCAAGAACTTAAACCTTCCAA 2940
 QY 826 PheGlnAlaAlaMetGlnProProAlaAspLysValAsnLeuGluProGlnAsnGlnAla 845
 DB 2941 TTTCAAGCTCCATGCAACCAACAGATTAAGCAAAATTTTAAGCAACAAACCAAGCA 3000
 QY 846 LeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyTyraThrAspGlyGluValAsp 865
 DB 3001 CTACGATTCAGGCTTATGATCTGTGGGATGTGTCAAGGTATCAACATATGGGAGGTCTGAT 3060
 QY 866 LysAsnGlnArgTyrglnSerThrLeuGlnPheAlaAlaAsnLeuLeuGlnIleGln 885

Db 3061 ABAATCAGAGATACCACTACATTTGGAGCAAACTTCTCCAGATACAG 3120
QY 886 GILyGInGInGInSerSerGlyThProAlGlyGInGlyPro 901
Db 3121 CAGAAACAGACAGACAGTCTTCCAGGTACCTCCGCTGACAGGGGCT 3168
RESULT 5
AX420419 4593 bp DNA linear PAT 18-JUN-2002
LOCUS Sequence 1 from Patent WO212518.
DEFINITION AX420419
ACCESSION AX420419
VERSION AX420419.1 GI:21524575
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1
AUTHORS Masino, R.M., Schomburg, F.M., Michaels, S.D. and Patton, D.
TITLES Floral induction gene
JOURNAL Patent: WO 0212518-A 1 14-FEB-2002;
WISCONSIN ALUMNI RESEARCH FOUNDATION (US)
FEATURES
source 1..4593
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
ORIGIN
Alignment Scores:
Pred. No.: 3,81e-167 Length: 4593
Score: 4392.00 Matches: 899
Percent Similarity: 58.72% Conservative: 0
Best Local Similarity: 58.72% Mismatches: 2
Query Match: 91.48% Indels: 631
DB: 6 Gaps: 4
US-09-920-705-3 (1-901) x AX420419 (1-4593)
QY 1 MetAlaIeuSerMetIysProPheArgAlaAspAspSerGlyPheGInSerAenAenLeu 20
Db 1 ATGGCGTTATCTTGAAGCATTCAAGCCGATGATTCGGTTTCCAGTCAACATCTT 60
QY 21 TTPValGlySerLeuThrProGluThrThGInSerAspLeuThrGluLeuPheGlyArg 40
Db 61 TGGGTCGGTAGCCTTAACGCCGAGACGACAGAGTCAAGATCTGACCGAGTTGTTGMAAG 120
QY 41 TTYGlyAspIleAspArgIleThrValTYrSerSerArgGlyPheAlaPheIleTYr 60
Db 121 TACGGGATATATATGATGAATCAACGGTGTATTTCTTCAAGAGCTTTCGGTTATATATCTAC 180
QY 61 ArgHisValGluGluAlaValAlaAlaIleGluAlaIleuGInGlyAlaAenLeuAenGly 80
Db 181 AGCAATGTGAGAGAGAGAGAGTCCGACCCCAAGAGGCTTCTCAAGAGCAATTTGAATGA 240
QY 81 SerGInIleIysIleGluTYrAlaArgPro----- 90
Db 241 AGTCAATTTAAGATCGAATACGACACCGGTTTGTCTTATCTATATCTTGGTTTGTTC 300
QY 90 ----- 90
Db 301 TCTAATTGATGTCTTTTGTCAACGATTATCTCTTTTGGAAATTCATAGTCAGGT 360
QY 90 ----- 90
Db 361 TCACAAAACCTTATGATGCTTGTATTAGTCCAAAATTTCTTGTGATCTGTTTTTTT 420
QY 90 ----- 90
Db 421 TTCCTATCGTACAAATCAAAAGTCGAAACCTAGTTTTTTTCTATATACGTCGTAGCTT 480

QY 90 ----- 90
Db 481 AAGGCGAAACCTGATCCGATCGAAACGCTTTTCTCAAAATTACTTTGTTATATGCACT 540
QY 90 ----- 90
Db 541 CGGCAAGACCAACACAGAGAGCTTCGAAAAATTTGATGTTAAAGCATATATACTC 600
QY 90 ----- 90
Db 601 TTAGCAATGAGCTCTGCAAAAAGAAATACATCAACAGATTTACTCTGTTATGCGAA 660
QY 90 ----- 90
Db 661 GAAAGTTAATCCGATTGCTGTTTATCTGCAAACTTGATCCGCTTACAGCTTGATT 720
QY 90 ----- 90
Db 721 TCAGTTCGCTTCAAGTGTAAATCTCAGACGCTTGAATGATGATATTTGGTACTCCA 780
QY 90 ----- 90
Db 781 CTCGAATTCGAATGTGAAATTTGAAAGAGCAACTCTATCATCGGCCAAACAGA 840
QY 90 ----- 90
Db 841 TAAAGAAATTTGAGATTAAAGTTCAAGTTCTCTGCAAAAATCAAAACCGGTGAGAAAT 900
QY 90 ----- 90
Db 901 TTGTCTATGCGGTACGTAGATATCAATACTGCATCTGCGAAACGAAATTTGGCAA 960
QY 90 ----- 90
Db 961 CCGTATCGTCTTAATCTTAATGATGCTGCGACATATCCATATCAAGTGAAGGAGTTT 1020
QY 90 ----- 90
Db 1021 GGGCTTGTGTTGGCTCGATGCTTCAAGAAACAAATTAAGTGTGTTAAGGGGCTAAC 1080
QY 90 ----- 90
Db 1081 TCTACCAATCAGAAACGCTTATTTGGAAGAACATGTTTGTCTCAATTCATCCATCCTA 1140
QY 90 ----- 90
Db 1141 CGTACATCTGGGCTTTCCCATTTAGTTCTTAGAAGTTGACTTCTTCAACACATTC 1200
QY 90 ----- 90
Db 1201 TGGGATTGTACAGTTGACAGTGAACAGACAAATATTCATTTGACAGCACTCAAAATT 1260
QY 90 ----- 90
Db 1261 CAATGCTTCCGGCTTTGAGTACTTAATGAGAGTTCAGGAGTTGTTATTTGGGAC 1320
QY 90 ----- 90
Db 1321 TCAGCAATGATTAAGACAGAAATGTTTGCAGAAAGTAAATATGCTTCCGCTTGG 1380
QY 90 ----- 90
Db 1381 GGGACTACTACTTCTCTACAGACAAATCTGATGTTTAGGCAAGAGAACTATTAACA 1440
QY 90 ----- 90
Db 1441 CTGATTTATAGAGAAAGAGAGATGTTTCTCTGGCGCAATTTTATTTGTTAGGAA 1500
QY 90 ----- 90
Db 1501 AGCAATGATATGAATGAGTGTGCTAGTGTAGTGAATTAATTAAGTTAGTTGTGTGTT 1560
QY 90 ----- 90

Db 1561 AGTTTCCTGATGTTGATGCTTATTCCTGGCAACCTATCTGGGTAGTATCGCCTTCT 1620
 QY 91 -----AlaIysProCyAllysSerLeuTPValGlyGlyIleGly 103
 Db 1621 ATCGACCTTTCTTGTGGCAGCAAAACCTTGTAAGTCTATGCGGTGGGGGAGATCGGC 1680
 QY 104 ProAsnValSerIysAspAspLeuGluGluIuphSerIysPheGlyLysIleGluAsp 123
 Db 1681 CCTATATGCTCCAAAGATGACCTGGAGGAAAGTTCACAGACTTTGGGAAATCGAGAT 1740
 QY 124 PheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
 Db 1741 TTTAGGTTTCTCAAGAGAACGACAGACAGCTTTCATTGATTATATAGATGATGATGCT 1800
 QY 144 LeuGluAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPhe 163
 Db 1801 TTACAGGCTAAGACATGATGAGAAAGCCTATGAGGTAGCTTTTGGCGTGTGATTTT 1860
 QY 164 LeuArgSerGluAlaProLysLys----- 171
 Db 1861 CTCGGGTCAACAGCGCCAAAAAGTAAGCACTTGTGGCATTTGATTTTACTTTTGA 1920
 QY 171 ----- 171
 Db 1921 AAAGCTCCAGTAACAATTTTGTGTTAGTTCAATATTTGGCTCAAACTGATAGGCTGAG 1980
 QY 171 ----- 171
 Db 1981 CTGTGCTTGCCCTCAGAGACAGATTTACTGCTGCTCTATTTCATTGTAGAGTAGGCT 2040
 QY 172 -----GluGlnTrpAlaGly 176
 Db 2041 CAACCTCTTAAGTCTGAATCAAGTTACCTTGTGTATCTTCAGAAACAATGGCTGGC 2100
 QY 177 SerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGln----- 190
 Db 2101 TCTTACATTAACAGAAATGGCAATGATCATTAACCCGA-GGTTAGTCTGAATGTTG 2159
 QY 190 ----- 190
 Db 2160 AAAGTATGCTCTGTACTAGTATGATGATAGTTACGGTTTGAAGTTTGTATAT 2219
 QY 191 -----TyrProHisSerTyrGlyAspPheLysGlyAspValGlnProSerLysVal 207
 Db 2220 TTTCTTACAGATCTCACTCATATGAAAGCTTTAAAGAGATGTCACAGCAAGTAAAGTT 2279
 QY 208 LeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsn 227
 Db 2280 CTGTGATTTGGGTTCCCTCTACTCTCACTCAACATGCAATGATGAGCAAAATCTGCACAT 2339
 QY 228 AlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPheAla 247
 Db 2340 GCGATATATCTCTTGTGTAGATCGAGAGGTAAAAAGTTACCAATCAAGAAATTTTGC 2399
 QY 248 LeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGlyLysGluGlnIleArg 267
 Db 2400 CTTGTGAGTTTAAAGCGCGAGAGAGGCTCCCAATGCAAGAAAGGCTTACAGGGAGG 2459
 QY 268 LeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProProGluGln 287
 Db 2460 TTATTCATATATCTAGATCAAAATTTATGTAATCAAAAGATGATTTGCTTCTGAGCA 2519
 QY 288 AspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAspAsp 307
 Db 2520 GACGATACAGATTGTTTACTCTGTATGAACCGGTCAAGACAGATATCTTCAATATGAT 2579
 QY 308 ProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgProLeuArg 327
 Db 2580 CCTTATTTGTATCTTCTTCTCTATCTATCTGGAATTTCTGAGTCTATAGAGCCCTCAGA 2639
 QY 328 GlyThrAsnGluArgSerTyrAsnGlyValaGluTyrAsnAspValaGlyLysGluPro 347

Db 2640 GGTACAAATGAGCGCTTCATATATATGTCAGAAATACATGACCTTGTGTAGAGCCA 2699
 QY 348 AsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGly 367
 Db 2700 AACTGAGAGAGCGCATCTGCAATATGAACTGGAATATCTCCATCTCCAAAGACCTGGA 2759
 QY 368 IleLeuProSerProAlaGlnGlyThrArgArgProMetLysSerAsnProAspSerTrp 387
 Db 2760 ATCTCTCCATCTCTGCAACAAAGTACAGAGCGCCCTTATGAGGTCAAAACCCGATCTTGG 2819
 QY 388 GluGlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgAspAspGlySer 407
 Db 2820 GAAGATATGATCTCTGCTCAGTTGTCAGAGAAATTAACAAACAGAGAGATGATGATCA 2879
 QY 408 ValAspGlyPheThrProMetGlyValaAspGluArgSerPheGlyArgLysValaAla 427
 Db 2880 GTGACGCTTTTACTCCATGAGGTGTGATGAGAGGTCAATTTGGTCGAGTTCAAGTTGCT 2939
 QY 428 AlaArgProIleArgGlyProProAspSerAspHisIleTrpArgGlyMetIleAlaLys 447
 Db 2940 GCTAAGCCTATCTCGTGGCCCTGATTTCTGATCATCATATGAGAGAAATGATGGCAAG 2999
 QY 448 GlyGlyThrProValCysCysAlaArgCysValaProMetGlyLysGlyIleGluThrLys 467
 Db 3000 GGTGAACCTCCGCTGTGTGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3059
 QY 468 Leu----- 468
 Db 3060 CTGTGAGTACTAATTTCTAGACACTTAAACCTTCTAGTGTCTTCTTTCAGAGCATTT 3119
 QY 468 ----- 468
 Db 3120 ATATATTTCCATTTCATTCTGCATGGAATACATTATTAATAGATAGTACATTTTATT 3179
 QY 468 ----- 468
 Db 3180 TTACTATTAATTGTTAGTTCTGAGATGCTGTGATTTTCATGATGTTTGTG 3239
 QY 469 -----ProGluVal 471
 Db 3240 GCATTGCCCTCAATTAAGCTTTGTTTTTTTAAATTAATGATTTATAGGCTGAGGT 3299
 QY 471 IValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisIleTyrAlaValAlaI 491
 Db 3300 CGTCAATTTGTTCAAGCAAGATGATTTGATATGCTGCTAAACATTAAGCCCTGTGCAT 3359
 QY 491 eGlyCysGluIleValaPhePheValaProAspArgGluGluAspPheAlaSerTyrThrG 511
 Db 3360 TGATATGAGATCTGTTTTTTTTCGTACACAGAGGAAAGATTTTGGCTTACACTGA 3419
 QY 511 uPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValaAlaLysLeuAspAspGlyTh 531
 Db 3420 ATTTCTCGGTCCTTATGCTCAAAAGATCGGGGGGTGGCCAAATTAAGATATGATGATC 3479
 QY 531 TrpLeuPheLeuValaProProSerAspPheLeuThrAspValaLeuGlnValaThrArg 551
 Db 3480 AACTTATTTCTGTGTGCTTCATCAATTTCTTAATGATGATCTCAAGTACCCCTGCA 3539
 QY 551 ngluArgLeuTyrGlyValaValaLeuLysLeuProProAlaValaProValaThrAlaSe 571
 Db 3540 AGAAGCGCTATATAGGTGTGTCTCAAGTTACCCCGCAGCGCTTCTGTATACAGCATC 3599
 QY 571 TrpTyrGlnGlnLysSerGlnSerAsnProLeuHisIleTyrMetAlaAspGlnAlaArgAspSer 591
 Db 3600 ATACAGACAAAGATCTCAAGTCCATCTCTGCAATTAATGATCAAGCCGGGATTCAC 3659
 QY 591 AlaAsnAlaSerHisSerLeuTyrProProArgGluAsnTyrIleArgValaProGly 611
 Db 3660 TGCCAAATGCTAGCACAGTTTATATCTCTCTAGGGAATTAATCATTAAGGGTGACAC 3719
 QY 611 uHisLeuThrAlaAlaSerLysProSerValaSerGluProLeuArgIleProAsnAsnAl 631
 Db 3720 ACATTTGACGCTGCTTCAAAACCATCTGTGTAGCGAGCTTCAGAAATACCTTAATATATGC 3779

QY 631 aaIaProGlnaIaGlyValSerLeuThrProGluLeuAlaThrLeuAlaSerIleLeu 651
 DB 3780 AACGGCTCAAGCTGGGGTAGTTAACTCCGAGCTTTTAAAGCACCATTATTTCT 3839
 QY 651 uProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetSerGlyProSerThr 671
 DB 3840 CCTTCGAATTCTCAACCTGCTGCTCCCTAGAGTCAACCAACCTATTCAGACCTTCAAC 3899
 QY 671 rValValSerThrAlaHisGlnSerAsnGlyLeuTyraGlnGlyValAlaProSerGlnAl 691
 DB 3900 ACTGTGTTTCCACAGCAGCATCAGTCCATGAGCATGTCAATGAGGAAGACCCGCTCAAC 3959
 QY 691 aThrPylAsaGlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTrpG 711
 DB 3960 TTGGAAAAAGAGGTCACAAACAGTTGATGATGCGTAATTCAGTCCATCCAAACAAATACG 4019
 QY 711 yAsnGlnTyThrProAlaGlyGlnLeuProProProProSerArgTyProProAlaSe 731
 DB 4020 AATTCAGTACACTCCAGCTGGGCACTACTCTCTCTGCGCTTACCTCCACTTCC 4079
 QY 731 rAnAsnProAsnTyThrSerGlyMetValHisGlyAsnMetGlnTyGlnSerGlnSe 751
 DB 4080 AAACAACCCCACTACACTAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 4139
 QY 751 rValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTySerMetTy 771
 DB 4140 TGTTAACAGAGCTCCAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4199
 QY 771 rThrGlnGlySerSerSerAsnHisProValSerGlnProMetValGlnGlnTyGlnProG 791
 DB 4200 CACTCGAGGTTCTCTCAATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4259
 QY 791 uAlaSerMetProAsnGlnAsnTyGlyProIleProSerTyGlnGlnAlaAsnDheH 811
 DB 4260 AACGTCATAGCCCAACCAAACTATGATGATGATGATGATGATGATGATGATGATGATG 4319
 QY 811 eGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetG 831
 DB 4320 TGGCGTAACAACTACAGGACAGAACTTAAACCTCTCCCAATTTCAAGCTCCATATGCA 4379
 QY 831 nProProAlaAspGlyAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMe 851
 DB 4380 ACCACACGACGATTAAGCAATTTAGAGCCAAACCAAGCACTACGATTCAGAGCTTAT 4439
 QY 851 tIleSerGlyAspGlyGlnGlyThrThrAspGlyGluValAspGlyAsnGlnArgTyG 871
 DB 4440 GATCTCTGGGAGATGTCAGGGTACAAAGATGGGAGGTGATTAAGATCAGAGATACCA 4499
 QY 871 nSerThrLeuGlnPheAlaAlaAsnLeuLeuGlnGlnIleGlnGlnIleGlnGlnGln 891
 DB 4500 GTCAACACATCAATTTGACGAAACCTTCTCTCCAGATTAACGAAACACAGCAACACA 4559
 QY 891 nSerSerGlyThrProAlaGlyGlnGlyPro 901
 DB 4560 GTCTTCAAGTACTCCGCTGACAGGGGCT 4590
 RESULT 6
 LOCUS AC002335/c 102057 bp DNA linear PLN 11-MAR-2002
 DEFINITION Arabidopsis thaliana chromosome 2 clone T1024 map C1C10F02,
 complete sequence.
 ACCESSION AC002335
 VERSION AC002335.3 GI:20196857
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (bases 1 to 102057)
 AUTHORS Rounaley,S.D., Techudy,M.M., Lin,X., Ketchum,K.A., Crosby,M.L.,
 Brandon,R.C., Spriggs,T.A., Mason,T.M., Kerlavage,A.R., Adams,M.D.,

JOURNAL Somerville,C.R. and Venter,J.C.
 REFERENCE 2 Unpublished
 AUTHORS Lin,X.
 TITLE Direct Submission
 JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 REFERENCE 3 (bases 1 to 102057)
 AUTHORS Town,C.D. and Kaul,S.
 TITLE Direct Submission
 JOURNAL Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA, cdtown@igr.org
 COMMENT On Apr 18, 2002 this sequence version replaced gi:598343.
 FEATURES
 source location/Qualifiers
 1..102057
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="2"
 /map="C1C10F02"
 /clone="T1024"
 1..16126
 /note="overlap with BAC clone F14B2
 (AC004450:74158..90283)."
 complement(334..355)
 /rpt_family="AT-rich"
 complement(3113..4397)
 /gene="At2g43270"
 /note="synonym: T1024.1"
 complement(join(<3113..3691,3734..3924,4004..>4397))
 /gene="At2g43270"
 complement(join(3113..3691,3734..3924,4004..4397))
 /gene="At2g43270"
 /codon_start=1
 /product="putative myb-related transcription factor"
 /protein_id="AAB64309.2"
 /db_xref="GI:20196861"
 /translation="MITYVDDLEIRIFLGLPIKSLIRFKTVSKOMRSILSKSPARR
 LNVKKKELIAYGDRTEIGFGESESIQMYTLHCDIDATRSITCBGLVCTPAPGMIV
 LNPSTRORRPFCSNNHVPDRIRFQFDELYLTSFGPNWMEGFRDFNSYKVR
 MCFSPVEKCEVLDVETGESEINPPNDIDVGRKSRFGILALDLTLERFNSVPSYC
 CTSQAMVNLDRLAITTKNGAEWELEILSLMDYDKELISKYSITSLASIIINHPWE
 SSWPYLVLTSLKGINVPCDNHRLPKYNTETNEICLSDISVISPYLELDIPLRLDS
 GHDHDKIRTSISCGLFSGHPEAGLFKEITTRFPATLVLYGVQYRSLKPK"
 5148..5767
 /gene="At2g43280"
 /note="synonym: T1024.2; predicted by geneScan"
 join(<5148..5447,5552..>5767)
 /gene="At2g43280"
 join(5148..5447,5552..5767)
 /gene="At2g43280"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAB64334.1"
 /db_xref="GI:2289005"
 /translation="MPTDDYSRRLGFPWAKVMSCRSEKDGRIARRFGCKGHCYSI
 RKGFSVRRPSPSTRREGCKAMTHVRSGRVITTKVEVHNHPLVSPREARHTIDE
 KDKRIOELTIELRNKRRLCAAYKEQIDAFKIVEHSNQIARKVENVVNNAKEFEHLE
 HALLRSKDAI"
 complement(5982..6953)
 /gene="At2g43290"
 /note="synonym: T1024.3; supported by full length cDNA:
 Ceres:31535"
 complement(<5982..>6953)
 /gene="At2g43290"
 complement(6214..6861)
 /gene="At2g43290"
 /codon_start=1
 /product="putative calcium binding protein"
 /protein_id="AAB64310.2"
 /db_xref="GI:20196862"
 /translation="WVRIRLIVNIINSFLSLVPPKRLRTLPFLSWFDPKTLHKNSPPSP


```

STMTPSSSSAPTAKRIDPSLKRIVQWFGNGGRIITKEELNDLENLGIATYIPDKDL
TOMIKIDANGDCVDDIDEPISYSIVDEHNDGEREEMKQAPVFPDDGGGPTT
VEELKSVASALGKOGKTLGCKKIMQVDDSGRVYKFEFLQMMKGSSSN"
repeat_region
complement(6998..7029)
/rpt_family="AT_rich"
repeat_region
complement(7486..7547)
/rpt_family="AT_rich"
repeat_region
complement(7628..7660)
/rpt_family="AT_rich"
gene
complement(10475..10550)
/gene="At2g43330"
complement(10479..10550)
/gene="At2g43300"
tRNA
complement(10479..10550)
/gene="At2g43300"
/product="tRNA-Thr"
11012..11413
/gene="At2g43310"
/notes="synonym: T1024.5; predicted by genescan"
<11012..11413
/gene="At2g43310"
11012..11413
/gene="At2g43310"
/codon_start=1
/product="hypothetical protein"
/db_xref="GI:2289004"
/translacion="MNGSTHRHLQLVLSCKRTAQVTPGSGSTIIAMASSEOEFLA
KSRANLYRPHSNHFDKTSRSGEGLGRLDELGVVSDADDEISRPVHHRKV
LPFDSVRRITGIVDGTGEQLNDISVPIIPKN"
complement(11649..14833)
/gene="At2g43320"
/notes="synonym: T1024.6; predicted by genescan; supported
by cDNA: g1_14517475_gb_AT039573.1"
complement(join(<11649..11950,12040..12519,12610..12654,
13002..13102,13191..13257,13742..13804,13936..14107,
14566..14833))
/gene="At2g43320"
complement(join(11780..11950,12040..12519,12610..12654,
13002..13102,13191..13257,13742..13804,13936..14064))
/gene="At2g43320"
/codon_start=1
/product="expressed protein"
/protein_id="AAB64311.2"
/db_xref="GI:20196858"
/translacion="MRAPSLAQCPLLPDRGVSALSEKDLQLPFAVEIIPSKT
VAHRRSGENLALGIQVFKGVADIIIGSGSTAPLKGSGKSSSVLVNVL
XNEIRDQGLSPFKRVLELGCNFGVGFACAGASVHFDLSAETTRCTTIPVLA
NLEQADRSQSPESPLPSROAISASVAFAGEBELSTVSIIRTVLEPSIPAMN
LSFSEDFMDGCSQDGSITQDPDSRSRSLSGSRAMERANETGSGEGYGVIIIM
TEIPYSITSLKLYSLIKKCLRPYGVWYLAKKQYGVFNGSAKHLRLVDEETLGA
HIVKETTRDIDIKFPLK"
16041..18948
/gene="At2g43330"
/notes="synonym: T1024.7"
join(<16041..16188,16408..16514,16563..16898,17700..17993,
18080..18948)
/gene="At2g43330"
join(16041..16188,16408..16514,16563..16898,17700..17993,
18080..18760)
/gene="At2g43330"
/codon_start=1
/product="putative membrane transporter"
/protein_id="AAB64332.1"
/db_xref="GI:2289003"
/translacion="WTLTIPNAPGSSGLDMPERRMSYFNSTYLLGLVTYAGIGLL
RGYDVGYSGLALTKDFEVVKOSFLQVNVSSFPSSKLETTVSALVQAMGAA
GAINVDYIKKATLEADVPFAAIIWAAPDPVILISGLILGVAVAPY
IAEASPSSEVGRGLVSTNVLMTGGFLSVNSAFQVGTWMLGVSGVPAVIOPI
LMFWPESRPLFMKNRKABAIQVLAITYDLSRLDELDHLSAEESEKORVGYL
DVFSKEIRLAFIAGAGAFQFOFGINTWVYSPTIYVAGSRSNGLAPLISITVA
MNAAGTVYGIYIDHCGKRLSLRQVITSLIISVFSKOSSTSDGLGWLVA
IGALITVFPAPGKPVPTVNSSETIPQIRIGICGMSATVNSLNIIVQETLLAE
AAGGTMFLIAGIATVAVIVFIVFPETQGLTPSEVEQIWKERAYNIGMSWSSDS
NNMEGLLEQSGS"

```

```

repeat_region
complement(21849..21900)
/rpt_family="AT_rich"
gene
complement(22675..23510)
/gene="At2g43340"
/notes="synonym: T1024.8"
join(<22675..23164,23243..23304,23325..23510)
/gene="At2g43340"
join(22675..23164,23243..23304,23325..23498)
/gene="At2g43340"
/codon_start=1
/product="unknown protein"
/protein_id="AAB64331.1"
/db_xref="GI:2289002"
/translacion="WCDSDFLRSGSYEADVDSDEISTSSPSSCSYSESEERINNG
FGVOSKQTKLEKKSNVLEGVVDSAVNDLKRKSLTDDELKGCVDLGF
FNVEISPELCOTLPALIELCYMSQKFIQDQHHSSSEPKSSVLSPPVPSRISWKT
SSPDNPDDVABRLKFMQAVACFMREFDDPKABESNEEEDETQFAGVSKMLRL
TCLKTTIVELPAHVTCIKRLMT"
complement(23549..25505)
/gene="At2g43350"
/notes="synonym: T1024.9; supported by full length cDNA:
Ceres:30737"
complement(join(23549..23749,24086..24133,24215..24382,
24465..24583,24673..24734,24830..24906,25278..25505))
/gene="At2g43350"
complement(join(24101..24133,24215..24382,24465..24583,
24673..24734,24830..24906,25278..25439))
/gene="At2g43350"
/codon_start=1
/product="putative glutathione peroxidase"
/protein_id="AAB64335.1"
/db_xref="GI:2289006"
/translacion="MPSRSRWVORATSKIKTILFGLVAFVLYLRYPSSPSTVEOS
STLYNISVDIDEGKDVLSLFTGKVLILVNVASKGLTEGNKEMILYAKYKQGF
EILAFPCNOSGDSNMEIKETVCNIFAESEPIFDKIEVNGKTCPLVPLKEQGF
GLFQDAIKMNFAPKILVDROGVVDRVAPTTSPLEIKDVKLASA"
complement(25700..28116)
/gene="At2g43360"
/notes="synonym: T1024.10; supported by full length cDNA:
Ceres:42038"

```

Alignment Scores:

```

Pred. No.: 1,032-165 Length: 102057
Score: 4392.00 Matches: 899
Percent Similarity: 58.72% Conservative: 0
Best Local Similarity: 58.72% Mismatches: 2
Query Match: 91.48% Indels: 631
DB: 8 Gaps: 4

```

US-09-920-705-3 (1-901) x AC002335 (1-102057)

```

QY 1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerIleuAsnLeu 20
DB 45895 ATGGCGTTATCTATGAGGCCATTCAGCGCAAGATTCGGTTTCCAGTCAAAACATCTT 45836
QY 21 TrpValGlySerLeuThrProGluThrTrpGluSerAspLeuThrGluLeuPheGlyArg 40
DB 45835 TGGGTCGATGCTTAACGCCGAGACACAGATCAGATCTGACCAAGTTGTTGGAAGA 45776
QY 41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyr 60
DB 45775 TACGGCGATATGATATGAAATCAACGGATTCCTTCACGAGCGTTGCGTTTATATATC 45716
QY 61 ArgHisValGluGluValAlaValAlaAlaValGluGluValAlaValGluGluVal 80
DB 45715 AGACATGTGAGAGAACAGATCGACCAAGAGCGCTTCAAGAGCAAAATTTGATAGA 45656
QY 81 SerGlnIleLysIleGluTyrAlaArgPro----- 90
DB 45655 AGTCAATTAAGATCAATACCAACAGACCGGTTGTTCTATCTATCTTCGTTTGTTC 45596
QY 90 ----- 90

```

Db 45595 TCTAATTGATGCTTTTTCACAGATTATACCTTTTGGAAATCATAGTCCAGGT 45536
 QY 90 ----- 90
 Db 45535 TCACAAACTTGTGATGCTGTGTTAGTCCAAAAATTTCTGTGAAATCTGTTTTTTT 45476
 QY 90 ----- 90
 Db 45475 TTCTCATGTACAAATCAAAAGTCAAACTAGTTTTTTCTATATATACGTGTTAGCTT 45416
 QY 90 ----- 90
 Db 45415 AAGCGAAACTGATCCGATTCGAAACGTCTTTTCTCAAAATTACTTGTATATCGAAGCT 45356
 QY 90 ----- 90
 Db 45355 CCGCGAAAGCCAAACACAGAGAGCTCTGCAAAATTTGATGTAAAGCATATATATATC 45296
 QY 90 ----- 90
 Db 45295 TTAGCAATGAGCTCTGCAAAAGATATACATCAAAACATGTTTACTCTGTTTATGCGAA 45236
 QY 90 ----- 90
 Db 45235 GAAGTTTATCGATGTCGTTTATCTGCAAACTTGATCCGCTTTAGAGCTGAT 45176
 QY 90 ----- 90
 Db 45175 TCAGTTTCGCTCAGATGTAAATCTCACAGCTTGATGATGATGTATGCGTACTCCA 45116
 QY 90 ----- 90
 Db 45115 CTCGAAATTCGCAATGCTGAAAAATTTGGAAGCGAGCAAACTTATCATGCGCCAAACAG 45056
 QY 90 ----- 90
 Db 45055 TAAGAAATTTTGAGATTAAAGTTCAGTTTCTGCAAAATCAACCGGTGAGAAAT 44996
 QY 90 ----- 90
 Db 44995 TTGCTATGCGGTACGTAGATATCAATATACATCTGCAAAACAGAAATTTCTGCGAA 44936
 QY 90 ----- 90
 Db 44935 CCGTATCGTCTAATCTTAATATGCTGCGACATATCCATATCAATGATGAGGAGTTT 44876
 QY 90 ----- 90
 Db 44875 GGGCTTGTGGCTGATGCTTACAGAAAGCAAAATTAAGTGTGTTAAGCGCGTAA 44816
 QY 90 ----- 90
 Db 44815 TCTACCAATCAGAAACGCTTATTTGAGAGAACATGTTGTCCTCAATTCCTCATCCCTA 44756
 QY 90 ----- 90
 Db 44755 CGTACATATGGGCTTTCCCATTTGATGTTCTTTAGGAAGTTGACTTTCAACAACATTC 44696
 QY 90 ----- 90
 Db 44695 TGGGAATGATACGTTGAGTGAAGCAAGCAAAATATTCAATTTGACAGCAGACTCAATTT 44636
 QY 90 ----- 90
 Db 44635 CATGTCTTCCGCTTTGAGTACTTAATAAGAGTGTGAGGATGTTTATTTGGGAC 44576
 QY 90 ----- 90
 Db 44575 TCAGCGAATGATAAAGACAGAAATTTTTCAGAAAGTATATGTCTTTCCGCGCTTG 44516
 QY 90 ----- 90
 Db 44515 GGGACTACATACTTCTCTACAGACAAATCTGATGTTTGAAGCAGAGAACTATTAACA 44456

QY 90 ----- 90
 Db 44455 CTGATTTATAGAGAAAGAGAGATGTTTCTCTGCGGCAATTTATTTGTTAGAA 44396
 QY 90 ----- 90
 Db 44395 AGCAATGATATGAAATGCTGTGCTAGTACTGAAATTAATTAAGTTTGTGTGTT 44336
 QY 90 ----- 90
 Db 44335 AGTTTCTTGATGTTGATGCTTTATTTCTTGCAACTATCTGGGATGATGCTCTT 44276
 QY 91 ----- 103
 Db 44275 ATGCACTTTTCTGTTGCGCAAAACCTTTGAGAGCTATAGGTGAGTGGGATCGG 44216
 QY 104 ProAmValSerIysAspSerLeuGluGluPheSerIysPheGlyIysIleGluAsp 123
 Db 44215 CCTAATGTCCTCAAGATGACCTGAGAGAGAGTTCAGCAACTTTGGGAAATTCAGAGAT 44156
 QY 124 PheArgPheLeuArgGluArgIysThrAlaPheIleAspTyrTyrGluMetAspAla 143
 Db 44155 TTAGGTTTCTCAGAGAACGCAAGACAGCTTCAATGATTAATAGAGATGATGATGCT 44096
 QY 144 LeuGlnAlaIysSerMetAsnGlyIysProMetGlyIysSerPheLeuArgIleAspPhe 163
 Db 44095 TTACAGGCTTAAAGCATGATGAAAGCCATAGGGGTGAGCTTTTGGGTGTGATTTT 44036
 QY 164 LeuArgSerGlnAlaProIysIys----- 171
 Db 44035 CTCGGTCAAGCCGCCAAAAAAGTAAGCACTCTTGCGCATTTGATTTTAACTTTGA 43976
 QY 171 ----- 171
 Db 43975 AAAGCTCCAGTAAACATTTTGTAGTTTCATAATTTGCGTCAAACTGATAGCGCTGAG 43916
 QY 171 ----- 171
 Db 43915 CTCTGCTTGTGCCCCCTAGAGACAGTATTAATCTGCTCTATTTGATGATGATGAGT 43856
 QY 172 ----- 176
 Db 43855 CAACCTTTAAGTCTGAATCAAGTACCTTTGTGTATCTTCAGAAACAAATGCGCTGCG 43796
 QY 177 SerTyrAspAsnArgAsnGlyIysMetAsnHisIysProGln----- 190
 Db 43795 TCTTACGATTAACAGAAATGCAATATGAAATCAATAACGCA-GGTTAGCTTGAATGTG 43737
 QY 190 ----- 190
 Db 43736 AAGATGCTCTGTTACTAGTATGATATAGGTACTGCTGTTTGAAGCTTTGTTATA 43677
 QY 191 ----- 207
 Db 43676 TTCTTACGATATCTTCACTCAATATGAAGCTTTAAAGAGATGTCCAGCAATAGATT 43617
 QY 208 LeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsn 227
 Db 43616 CTGTGGAATGGGTTCCCTCTCTACTGCTCAACATCAATGATGACAAATTCGCACAT 43557
 QY 228 AlaMetIleLeuPheGlyGluIleGluArgValIysSerTyrProSerArgAsnPheAla 247
 Db 43556 GCGATGATATCTTTGTGGAGATTCAGAGGGGTAAGTTAAAGTCCATCAAGGAATTTTGA 43497
 QY 248 LeuValIleuPheArgSerAlaGluGluAlaArgIleCysIysGluGlyLeuGlnGlyArg 267
 Db 43496 CTTGTGAGTTTAGAGCGCGAGAGAGCTCCCAATGCAAGAGAGGCTTACAGGGGAG 43437
 QY 268 LeuPheAsnAsnProArgIleIysIleMetTyrSerAsnAspGluLeuProProGluGln 287
 Db 43436 TTATTCATTAATCTAGATCAAAATTAATATGATCAAAAGATGAGTGTGCTCTGAGCA 43377

QY 288 AspAspThrSerPheTySerGlyMetLeuArgSerArgThrAspMetPheAsnAsp 307
DB 43376 GAGCATCTAGTCTTCTGATGTAACCGGCAAGCAAGTATGTTCAATATGAT 43317
QY 308 ProSerCyValSerSerProHisSerThrGlyIleProGlySerMetArgProLeuArg 327
DB 43316 CCTCATTTTATCTCTCTCATTTCTAGTAATCTGGGCTATAGAGCCCTCAGA 43257
QY 328 GlyThrAsnGlyLeuArgSerTyraAsnGlyValGluTyraAspValValGlyArgGluPro 347
DB 43256 GGTACAAATAGAGGTTCAATATATGTCAGAAATACATGACCTGTGTAGAGGCCA 43197
QY 348 AsnTPraArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGly 367
DB 43196 AACTGAGAGAGGCATCTGCATATGAAATCGAATCTCCATCTCCACAGACCTTGA 43137
QY 368 IleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTP 387
DB 43136 ATCTCCCATCTCTCTGCACAAAGGTACGAGCGCCCTATGAGTCAAAACCCGATCTTG 43077
QY 388 GluGlyTyraAspProAlaGlnLeuValArgGluSerLeuArgThrArgArgAspGlySer 407
DB 43076 GAAGCATATGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 43017
QY 408 ValAspGlyPheThrProMetGlyValAspGlyArgSerPheGlyArgGlySerValAla 427
DB 43016 GTGAGCGGTTTACTCCATGGGTGTCATGAGAGGTCAATTTGGTCAAGTTTCAAGTTGCT 42957
QY 428 AlaArgProIleArgGlyProProAspSerAspHisIleTPraArgGlyMetIleAlaArg 447
DB 42956 GCTAGACCTATCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 42897
QY 448 GlyGlyThrProValCysCysAlaArgCysValProMetGlyLeuArgGlyIleGluThrArg 467
DB 42896 GGTGGAACCTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 42837
QY 468 Leu----- 468
DB 42836 CTGTGAGTACTAATTTCTAGCACTTAAACCTTCTAGTGTCTTTTCAAGCGATTT 42777
QY 468 ----- 468
DB 42776 ATATATTTCCATTTCTATCTCTGAGTGAAGTAACTATTAATGATGATGATTTTAT 42717
QY 468 ----- 468
DB 42716 TTACTATTACTTGTAGTTTCTGAGATGCTTGATTTTCATGGTTCATTTTTC 42657
QY 469 -----ProGluVal 471
DB 42656 GCATGCCCCCTCAATTACTGACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 42597
QY 471 lValAsnCySerAlaArgThrAspLeuAsnMetLeuAlaValSHISlyrAlaValAlaIle 491
DB 42596 CTTCAATTTCTGACGAAGACTGATTTGATATGCTCGCTAAACATACGCGGTTCAT 42537
QY 491 eGlyCysGluIleValPhePheValProAspArgGluLeuAspPheAlaSerTyThrGly 511
DB 42536 TCGAGTGTGAGATCGTTTGT 42477
QY 511 uPheLeuArgTyrlLeuSerSerLeuAspArgAlaGlyValAlaValLeuAspAspGlyTyh 531
DB 42476 ATTTCTCCGGTACCTTAAGCTCAAAAGATCGGGGTGTGTGCAAAATTAATGATGTGAC 42417
QY 531 rThrLeuPheLeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGly 551
DB 42416 AACTTATTTCTTGAGGCTCCATCAGATTTCTTAATCTATGATCTCCAAAGTACCCGCTCA 42357
QY 551 nGluArgLeuTyrlGlyValValLeuLeuLeuProProProAlaValProValThrAlaSer 571
DB 42356 AGAAGGCTATATGT 42297
QY 571 rTyrlArgGlnGlnSerGlnSerAsnProLeuHisTyrlMetAspGlnAlaValAspSerPr 591

DB 42296 ATATGACACAGAACTCTGATCAATCTCTGTATTAATGATGATCAAGCCGGGATTCACC 42237
QY 591 oAlaAsnAlaSerHisSerLeuTyrlProProArgGluAsnTyrlIleArgGlyAlaProGly 611
DB 42236 TGGCATGTGATGACAGTTTATATCTCTTAAGGGAATTAATTAATTAAGGGGTGACCA 42177
QY 611 uHisLeuThrAlaAlaSerTyrlProSerValSerGluProLeuArgIleProAsnAla 631
DB 42176 ACATTTGACAGTGTCTCAAAACCATCTGTATGACGAGCTCTCAGAAATTAATATG 42117
QY 631 aAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThrLeuAlaSerIleLe 651
DB 42116 AGCGCTCAAGCTGGGGTTAGTTTACTCCGAGCTTTTACCACTGTGGATTCATTTCT 42057
QY 651 uProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetSerGlyProSerTh 671
DB 42056 CCTTCACACTTCTCAACTGCTGCGCTGAGAGTACACCACTATGTCAGGACCTTCAAC 41997
QY 671 rValValSerThrAlaHisGlnSerAsnGlyLeuTyrlAsnGlyValAlaProSerGlnAl 691
DB 41996 AGTTGTTTCCACAGACATCTGATCCATGAGCTGTATCAATGAGAAAGACCGTCTCAAG 41937
QY 691 aTPraArgGlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyrlGly 711
DB 41936 TTGGAAGAGAGTCCACAAACATTCATGATGCTCAATATGATCATTCACAAATACGG 41877
QY 711 yAsnGlnTyrlThrProAlaGlyGlnLeuProProProProSerArgTyrlProProAlaSe 731
DB 41876 AAATCAGTACACTCAGCTGAGGCAACTACCTCTCTCTGCGGTATCCCTCAGCTTC 41817
QY 731 rAsnAsnProAsnTyrlThrSerGlyMetValHisGlyAsnMetGlnTyrlGlnSerGlnSe 751
DB 41816 AAACAAACCCCACTACACTATGAGAAATGTCATAGGCAACATGAAATACAGAGCCATC 41757
QY 751 rValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyrlSerMetTy 771
DB 41756 TGTTAACATGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 41697
QY 771 rThrGlnGlySerSerAsnHisProValSerGlnProMetValGlnGlnTyrlGlnProGly 791
DB 41696 CACTCAGGTTCTGCAATATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 41637
QY 791 uAlaSerMetProAsnGlnAsnTyrlGlyProIleProSerTyrlGlnGlnAlaAsnPheHis 811
DB 41636 AGCTTCATGCTCAACCAAACTATGCTCCATTCCAAGTTATCAGCAAGCTTAATTTCA 41577
QY 811 eGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGly 831
DB 41576 TGGCGTAAACAATCAGGACAGACGAACTTAACCTTCCATTTCAAGCTGCATGCA 41517
QY 831 nProProAlaAspValAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMe 851
DB 41516 ACCACCAAGCATTAAGGCAATTAAGGACCAAAACCAAGCACTACGATTTGAGCTAT 41457
QY 851 tIleSerGlyAspGlyGlnGlyThrThrAspGlyGlyValAspLeuAsnGlnAlaArgTyrlGly 871
DB 41456 GATCTCTGGGAGATGCTAGGCTACACAGATGGGAGGTGATGATGATGATGATGATGAT 41397
QY 871 nSerThrLeuGlnPheAlaAlaAsnLeuLeuGlnIleGlnGlnLeuArgGlnGlnGln 891
DB 41396 GTCAACACTCAATTTGACCAAAACCTTCTTCCAGATGACAGCAAGAACAGACGAACA 41337
QY 891 nSerSerGlyThrProAlaGlyGlnGlyPro 901
DB 41336 GTCTTCAGTATCTCGCTGAGACAGGGGCT 41306
RESULT 7
AX420422/c 801 bp DNA 1lineat PAT 18-JUN-2002
LOCUS AX420422
DEFINITION Sequence 4 from Patent W00212518.
ACCESSION AX420422
VERSION AX420422.1 GI:21524578

FEATURES
source

CDS

mRNA

```
complement(22328, .22807)
```



```

/note="predicted by GeneMark_hmm etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BA033819.1"
/db_xref="GI:50726283"
/translation="MPAPASSRRVPLAAPAEQRHDEAEBCCYAATTTTRVYVPLRRRG:

```

Alignment Scores:	
Pred. No.:	1,63e-34
Score:	1117.50
Percent Similarity:	45.76%
Best Local Similarity:	32.40%
Query Match:	23.28%
DB:	8
Gaps:	39
Length:	124083
Matches:	325
Conservative:	134
Mismatches:	280
Indels:	264
Gaps:	39

US-09-920-705-3 (1-901) x AP005735 (1-124083)

[illegible]

OY	300	rgthrAspMetPheAsnAenApProSerCyValSerProHisSerThrnglyIleP	320
Dd	41832	-----GAGAGTATTATGATGATGTCTAGTAGTGCCCTCTGATTA	41791
OY	320	roglYserMetArgProLeuAargLYThraSngluArgSerTYraAngLYlaGIuTYra	340
Dd	41790	-----TTTGATTCCTCGACGGGGGAAGAATCAAGATAC-----TTTGAAATAACA	41751
OY	340	snaePvalValGLyLSgLUProAsnTPraRgaRProSeRaLaengLYthRngLYleL	360
Dd	41750	GTCGTGC-----	41743
OY	360	eUProSerProThrngLYProGLYtleuPProSerProAlaengLYthRaRgaPrOM	380
Dd	41742	-----CCTGTTTCTGCTGTGATTTCTTCATCACCGGAATCTGGA-----AAATCCC	41697
OY	380	eTaRgaSenAProAbPseRTripoluGLYtyTraSPProAlaengLeuValaRguISerL	400
Dd	41696	TCTTAACGTGACGATCTGCTCAGAGTACATTCGATCCA-----AGAGAAGCCA	41649
OY	400	ySaRTrArGaRgaepSLySerValASP-----GLYP	411
Dd	41648	AGAGSTTGCCTGGTGGATGCTGCACAGCTGACCCTTAATATACAAAGGACGTAAGAGGTC	41588
OY	411	heThrPrometGLyValaSpGLuaRgSerPheGLYarGLySerValaLaalaaRgPro	430
Dd	41588	TTTATTCTGCTGGGTACTCTCACGCGGAGATGACGTTCTGAGAGAGCATCAAGTCCTG	41528
OY	431	-----leaRgLYProProAbPseR-----AphsileTRpaRgLYM	444
Dd	41528	CTATTTCGAATTCATGGACAGTGCACCGAACATCATACCTTGAGCACTTGGCGGTGTA	41465
OY	444	ecIleAlaLySGLYgLYthProVALcyeySaIlAaRgySaValPrometGLYyGLYI	464
Dd	41468	GCAITGGCAAAGGTGGATCTCTGTATGTGTGCTGCTGCTTTGCCCTTAAGAAGAGGTG	41400
OY	464	legIuthRyLSU-----	468
Dd	41408	TTGAGATACCTTGTATGTATACATCACCATCTTAATCTTGTTCTCAAGTGAATTC	41348
OY	469	-----ProGUValValaenCYs	475
Dd	41348	TATAGCCTCGATCTTACACATATTCACCTGTTTTGCAGACCGAAGTGTAAATGTT	41289
OY	475	exAlaAcGYThraPLeuAsnMetLeuAlaLySIstYraLaVaLaIleGLYcyGuRI	495
Dd	41288	CCGCTTAACCTGCACTAGATATCTGTGCMAAGCATTAATAGACGCTTCAGGCTTGATA	41228
OY	495	leValPhePheValProaspARGLUGLuabSPheAlaSerTYThRngluPheLeuarGT	515
Dd	41228	TTCGCTCTCTCTTSCCAGATAGAGAAATGACTTTCTTTCTTAACATGAATTTTTCGCT	41168
OY	515	yrIeuSerSerLYsaBPARGAlaGLYValaLaLyLeuabSPbpgLYThRThLeuPheL	535
Dd	41168	ACTTAGGCTCAAAAAGTCGGGACGGGGTGTAAAGGTATAGGGGGGACTACTTAATTT	41109
OY	535	eUValaProProSarPePheLeuThraSPyaLleuengLYThRaRgIngiuATgLEut	555
Dd	41108	TGGTCCCAACCTCCGATTTTTCAGAAATGTTTGCAGATGATGCTCCAGAAAGCTTT	41048
OY	555	yrGLYValValleuLYsleuPro-----ProProAlaValProVALT	569
Dd	41048	ACGGTGTAGATTNACATTCGCCCAATGTCTGCTCTGCTCTGCTTCGCTCCACATC	40989
OY	569	hralaSerTYraRgInglu-----SerGlnSerAnProLeuHISTryMeLaBpg	586
Dd	40988	CTGCTGTACAAAGGCCCAACTAATGCACCAAGAAACAACAACCT-----TACTACGAG	40935
OY	586	InAlaAgPaSerProAlaSnAlaSerHIserIEuRYrProARagLIuaanTYri	606
Dd	40934	AA---AGGAAATTCCTTTGCAAGAAAGTATGTATGATTACGCTTACCAACAATCAC	40878
OY	606	leaRgGLYalaProGLUHIsleuThRaLa-----	615

```

Db      40877 ATCGAGATGCT---GATCAGCGTGAGTCTTACGTGAGAGATTCATGATCATTGGGGC 40821
Qy      616 -----AlaserlyProserValserGluPro-----LeuArgI 627
Db      40820 AAATCTTTCGACGACCTCGGTGGATGATGAGCAAGTAGTTCAGCCAAACCTTGCCGGTA 40761
Qy      627 leProAsenAlaAlaProGlnAlaGlyValSerLeuThrProGlnLeuLeuAlaThrL 647
Db      40760 TCCCTACAAAGCAGAGACTACAGCTACAGCTTACCTCCAGCTGATATATAGCCACTT 40701
Qy      647 euAlaserlLeuProAlaThrSerGlnProAlaAlaProGlnSerHISGlnPromets 667
Db      40700 TAGCTAAACTTTTCCCAAGTGGGAGCATCAGCTCGGCACCTGGCCACTTCCCTGTA 40641
Qy      667 erglyProserThValValSerThzAlaHISGlnSerAnGlyLeuThyAsnGlyGlu 687
Db      40640 GC-----TCAAAGCAGCAACCTGCATTAACCCAAATGAAT-----GATGCACTTA 40596
Qy      687 laProserGlnAlaThrLysArgGlyProGlnThrValHISAspAlaserAnGlnSerP 707
Db      40595 CTCCTTGCAAAAGTATGAGAG-----CTGAAATACAGCAATGCTTCTTACTTCACTCT 40542
Qy      707 heGlnGlnThyGlyAsnGlnThyThProAlaGlyGlnLeuProProProProSerArgT 727
Db      40541 TAGAGCAGATCGTAAT---TTCAGCATTCAGGACAA-----CAGTTACAGCAAC 40494
Qy      727 yrProAlaserAnAsnProAsnThyThSerGlyMetValHISGlyAsnMetGlnT 747
Db      40493 AACCTGAGAGCATTCATCTGCCAAATAT-----GGAACTTGGAGGAGTGACAGAG 40440
Qy      747 yrGlnSerGln-----SerValAsnMetProGlnLeuSerP 759
Db      40439 ACCCAACACAAACATAGTGTTCATCATCTGAAATGATGATCTAAACTGCCACACACCCAC 40380
Qy      759 roLeuProAsnMetProHISAsnThySerMetThyThGlnGlySerSerAnHISerP 779
Db      40379 CACCTCTTACGCTGCCACCCAGTTCGTATATATATACATCGCAAGTGGACACAGTTTGC 40320
Qy      779 roValSerGlnProMetValGlnGlnThyGlnProGluAla---SerMetProAsnGlnA 798
Db      40319 CT---ACGAGAGATGATGACAAACATATACACAGAGAGATTTATATAGACTCAAGCA 40263
Qy      798 snThyGlyProIleProSerThyGlnGlnAlaAsn-PheHISGlyVal----- 813
Db      40262 ATTATGTCATATGGCCACAGTTAGCTTAATCTTCAAGTTCAGCAATCTTAATCC 40203
Qy      814 -----ThrThAsnGlnAlaGlnAsnLeuAsn 822
Db      40202 CATTCTTCTTATTCCTCAAGTAACCTTGCACAAATATATCAATGAGGTTTTCG--- 40146
Qy      823 ProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsn----- 838
Db      40145 ---GCAGTGTTCAGAGCTCTTGACTTGGCTTAATGATAGGTAACATATGCTTCT 40089
Qy      839 LeuGlnProGlnAsnGlnAlaLeuThrGluGlnProMetLieserGlyAsp----- 855
Db      40088 TTACAGGCGCACATCAACAAATGTGTGCACTCTGTGCACAACTCTGTGTCGCG 40029
Qy      856 -----GlyGlnGly 858
Db      40028 CAATTCCTCTCGTATGCAAGCTCTGTCGCCGACAAAGCTCTGTGTCGACAAAGCT 39969
Qy      859 ThrThAspGlyGlyValAspLysAsnGlnArgThyGlnSerThrLeuGlnPheAlaAla 878
Db      39968 TCTGTGTGAC---GAGGCAAGACGAAACAGAAAGTATCAGGGAAGCTCTCAATGGCTAG 39912
Qy      879 AsnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 898
Db      39911 CGCTGTTGGTGCACATCAACAGAAAGCTGGAATCAACTTAAGGCTATCTCGAAAT 39852
Qy      899 Gln 899

```

```

Db      39851 CAA 39849
RESULT 10
ACI08755/c
LOCUS
DEFINITION
Oriza sativa (japonica cultivar-group) chromosome 9 BAC clone
OSJNBa0028C22, complete sequence.
ACI08755
VERSION
ACI08755.2 GI:50345123
KEYWORDS
SOURCE
ORGANISM
Oriza sativa (japonica cultivar-group)
Oriza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzoae; Oryza.
1 (bases 1 to 154228)
REFERENCE
Lee,J.-S., Hahn,J.-H., Yoon,U.-H., Lee,K.-S., Kim,Y.-H., Eun,M.Y.,
Kim,H.-I., McCombie,W.R., Zutvern,T., de la Bastide,M.,
Spiegel,L., Muller,S., Nascimento,L., Balija,V., Bell,M.,
Miller,B., Katzenberger,F., Andrade,M.V., Dike,S., O'Shaughnessy,A.
and Palmer,L.
Oriza sativa (japonica cultivar-group) chromosome 9 BAC clone
OSJNBa0028C22, complete sequence
2 (bases 1 to 154228)
REFERENCE
Hahn,J.-H. and Kim,H.-I.
Direct Submission
Submitted (31-JAN-2002) Rice Genome Sequencing Project, National
Institute of Agricultural Science and Technology(NIAST), RDA, 249
Seodun-dong, Suwon 441-707, Korea (E-mail:jhhahn@rda.go.kr,
Tel:82-31-290-0309, Fax:82-31-290-0308)
3 (bases 1 to 154228)
REFERENCE
Hahn,J.-H. and Kim,H.-I.
Direct Submission
Submitted (16-JUL-2004) Rice Genome Sequencing Project, National
Institute of Agricultural Biotechnology (NIAB), RDA, 225
Seodun-dong, Suwon 441-707, Korea
COMMENT
On Jul 16, 2004 this sequence version replaced gi:18449960.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. OSJNBa0028C22
overlaps clone OSJNBa0065A15 (ACI37594) from base 1 to base 12997.
The overlap is from base 155808 to base 168804 on OSJNBa0065A15.
OSJNBa0028C22 also overlaps clone OSJNB0034B12 (AP005735) from
base 102828 to base 154228. The overlap is from base 1 to base
51401 on OSJNB0034B12. This BAC clone was sequenced to phase II by
the Korean National Institute of Agricultural Biotechnology and was
completed to phase III by the Cold Spring Harbor Laboratory Genome
Center.
FEATURES
source
location/Qualifiers
1..154228
/organism="Oriza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="9"
/clone="OSJNBa0028C22"
/clone_1kb="HindIII"
56468..56601
/note="The assembly is covered by high quality sequences
derived from a transposed plasmid subclone."
ORIGIN
Alignment Scores:
Pred. No.: 2,06e-34 Length: 154228
Score: 117.50 Matches: 325
Percent Similarity: 45.76% Conservative: 114
Best Local Similarity: 32.40% Mismatches: 280
Query Match: 23.28% Indels: 264

```



```

Db      143527 TAGTAAACTTTGGCAAGTGGGAGCATGACGCTGGTCACTGGCCAGCTCCCTCGA 143468
Oy      667  englyProserThrvAlvalserThrvAlahisginSeranglyLeuThrvanglyGUA 687
Db      143467 GC-----TCAACGAGCAACACCTGATTACCCAAATGAAT-----GATGCATCTA 143423
Oy      687  IaproSerGlnAlaThrvAlvalserThrvAlahisginSeranglyLeuThrvanglyGUA 707
Db      143422 CTCCTGGCAAAAGTATGAGA-----CCTGAATAATCAAGCAATGGCTTTTACTTCACTCT 143369
Oy      707  heginGlnThrvAlvalserThrvAlahisginSeranglyLeuThrvanglyGUA 727
Db      143368 TAGGAGCATGCTGGTAAAT-----TTCCAGCATTCAGAGCAAA-----CAGTTCAGCAAA 143321
Oy      727  YrProProAlaSerAsnAsnProAntyThrvThrvserGlyMetValHisGlyAsnMetGlnT 747
Db      143320 AAGCTGAGAGCATGCTCACTCTGCAAACTAT-----GGAACTTGGCAGGTCACAGAGGC 143267
Oy      747  YrGlnSerGln-----SerValAsnMetProGlnLeuSerP 759
Db      143266 ACCCAACACAAACATAGTGTTCATCAATCCTGAAATGACTTAATCTTCCACACACACAC 143207
Oy      759  roLeuProAsnMetProHisAsnAsnThrvserMetThrvGlnGlySerSerAsnHisP 779
Db      143206 CACCTCTAGCGTCCACCCAGTTCCTATATTATATCATGCAAGTTGGACACAGCTTTC 143147
Oy      779  roValSerGlnProMetValGlnGlnThrvGlnProGlnAla-----SerMetProAsnGlnA 798
Db      143146 CT---ACGAGATAGTACGACACACACACACACACACACACACACACACACACACACAC 143090
Oy      798  enThrvGlyProAlaProserThrvGlnGlnAlaAsn-PheHisGlyVal----- 813
Db      143089 ATTATGTCTCAATTGGCCACAGTTAGCAGTTCTATCTTCACTGACCAATTTCAATCC 143030
Oy      814  -----ThrvAsnGlnAlaGlnAsnLeuAsn 822
Db      143029 CATTCTCTCTTATCCCAAGTAACTTCCCAACAAATATATAGAGGCTTTTGG--- 142973
Oy      823  ProSerGlnPheGlnAlaAlaMetGlnProProAlaAspPheValAsn----- 838
Db      142972 ---GAGTGTTCAGCATGCTTTCAGTTCCTTAATGATAGGTAACATATGCTTCTCT 142916
Oy      839  LeuGlnProGlnAsnGlnAlaAlaMetGlnPheMetCileSerGlyAsp----- 855
Db      142915 TTAAGGCGCACCATCAACAAATTTGTCGATCTCTGCAACAGCTCTGTCGTGG 142856
Oy      856  -----GlyGlnGly 858
Db      142855 CAATCCCTCTGCTATGCAAGCTCTCTGCGCCGCAACAGCTCTGTCGACACAGCT 142796
Oy      859  ThrvThrvAspGlyGlnValAspPheValAsnGlnThrvGlnSerThrvLeuGlnPheAla 878
Db      142795 TCGCTGAC---GAGGACAGAACGAGAACAGAGTATCAGGCACTTTCAGTTGCTCG 142739
Oy      879  AsnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 898
Db      142738 CGCTCTTGGCTAGCTACAAAGAGCCTGGAATCAACTTAAGGCTATCTGAAAT 142679
Oy      899  Gln 899
Db      142678 CAA 142676

```

```

RESULT 11
AP006756/1 176553 bp DNA linear PLN 28-JUL-2004
LOCUS      Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 9,
DEFINITION PAC clone: P0450E05.
ACCESSION  AP006756
VERSION     AP006756.2 GI:50725340
KEYWORDS   Oryza sativa (japonica cultivar-group)
SOURCE     Oryza sativa (japonica cultivar-group)
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

COMMENT

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

1
Sasaki, T., Matsumoto, T. and Katayose, Y.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, PAC clone: P0450E05
Published Only in Database (2004)
2 (bases 1 to 176553)
Sasaki, T., Matsumoto, T. and Katayose, Y.
Submitted (31-MAR-2004) Takuji Sasaki, National Institute of Agricultural Sciences, Rice Genome Research Program, Kamondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@nias.affrc.go.jp, URL: http://exp.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468)
On Jul 27, 2004 this sequence version replaced gi:46093787.
Genes were predicted from the integrated results of the following:
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH (http://www.softberry.com/), GeneMark.hmm (http://opal.biology.gatech.edu/geneMark/), GlimmerX (http://www.tigr.org/tdb/glimmer/gimr_form.html), RiceHMM (http://exp.dna.affrc.go.jp/RiceHMM/), SplicePredictor (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The genomic sequence was searched against NCBI Nonredundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DDBJ. Protein homologues of the coding regions were searched against NCBI Nonredundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no.
A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.
The orientation of the sequence is from 77 to 596 of the PAC clone. This sequence of P0450E05 has an overlap with OSUNB0055A15 (DDBJ: AC137594) clone at 5' end and with OSUNB0034812 (DDBJ: AP005735) clone at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://exp.dna.affrc.go.jp/GenomeSeq.html>.

FEATURES

gene
mRNA
CDS

```

1. 176553
  /organism="Oryza sativa (japonica cultivar-group)"
  /mol_type="genomic DNA"
  /cultivar="Nipponbare"
  /db_xref="taxon:39947"
  /chromosome="9"
  /clone="P0450E05"
  1479..4246
  /gene="P0450E05.1"
  join(<1479..1885,3329..3366,3854..3945,4100..>4246)
  /note="start and end point are not identified"
  join(1479..1885,3329..3366,3854..3945,4100..4246)
  /gene="P0450E05.1"
  /note="similar to Arabidopsis thaliana chromosome 5,
  At5g47870"
  /codon_start=1
  /product="hypothetical protein"
  /protein_id="BAD34413.1"
  /db_xref="GI:50725341"
  /translation="MEALRLSGPAAPPEVPAARVAAMGARRRGRSSSVVAAMEG
  GGCGKGVPTTNYVPLDKATGMTPLVEITRLDNLNRVDKIIDPTNTVPWSLPLIP

```

[illegible][illegible]

Oy		759	Proteinase	mechE	ProHisa	asna	SerTyrSerMetCysLeuGlnGlySerSerAsnHisP	779
Db		161625	CACCTCCCTAGCGTCGCCACCACCACTGTCTGCATATATTATCATCTGGCAAGTTGGACAACAAGTTTGC					161566
Oy		779	rovalserGlnPrometvalGingIntYrGlnProGlnala---	SerMetProAsnGlna				798
Db		161565	CT---ACGCAGATGATGATCGAAGCAAGTATCAACAGAGCAGTAGTATATGACTCAAGAACA					161509
Oy		798	enTYrGLYProIleProSerTyrTGInGlnAlaasn-PheHiSgLYal-					813
Db		161508	ATTATGTCATATTTGGCCACAGATTGAAGCATTTTAATCTTCAAGTCAGCAATTTCTAACATCC					161449
Oy		814	-----ThcThcAsnGlnlaIeGlnAsnLeuAsn					822
Db		161448	CATTCTTCCTTATTCCTCCCAAGTAACCTTGCCACAAATATATAGATAGGGTTTTGG---					161392
Oy		823	ProSerGlnPheGlnAlaIaIametGlnProProAlaAspLysAlaasn-----					838
Db		161391	--GCAGTGTTCAGCATCTCTTTGACTTGTGCTTAATGATAGGGTAAACTATGATGTTCTCT					161335
Oy		839	LeuGIuPProGlnAsnGlnAlaIeuatyleuatyLeuGlnPrometIleSerelyasp-----					855
Db		161334	TTCACAGCGCACCATTCMAAATAATGTTGGCACTCTGCTGCACACAAGCTCCTGTTGGTCGC					161275
Oy		856	-----GlyGlnGly					858
Db		161274	CAATTCCTCCTCGTCATATGCAGACTCCTGCTGCCGACCAAGCTCTGTTGCTGCACAAAGCT					161215
Oy		859	ThrThrAspGlyGluValaAspLysaenglnaryTyrglnSerThrLeuGlnpheAlaIa					878
Db		161214	TCTGCTGCAC---GAGCGAAGAACGGAACAGAAATATCAGGCGCACTCTTCAGTTGGCTTCAG					161158
Oy		879	AseLeuLeuLeuGlnIleGlnGlnLysGlnGlnGlnSerSerGlyThrProAlaGly					898
Db		161157	GCGCTGTGGGTGCTGACTCAACAGAAAGCCTGGAATATCAACCTTAAGGCTATCTCTTGAAAT					161098
Oy		899	Gln					899
Db		161097	CAA					161095
RESULT 12								
AK100889		3455 bp	mRNA					linear
LOCUS								PLN 24-JUL-2003
DEFINITION			Oryza sativa (japonica cultivar-group)					cDNA clone:J023129A05, full insert sequence.
ACCESSION		AK100889						
VERSION		AK100889.1						GI:32986098
KEYWORDS		FLI cDNA; CAP trapper.						
SOURCE		Oryza sativa (japonica cultivar-group)						
ORGANISM		Oryza sativa (japonica cultivar-group)						
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;						
		Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;						
		Ethnariolideae; Oryzeae; Oryza.						
REFERENCE		1						
AUTHORS		The Rice Full-length cDNA Consortium, National Institute of						
		Agricultural Sciences Rice Full-length cDNA Project Team,						
		Kiruchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,						
		Kishimoto,N., Yezaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,						
		Ohtsuka,K., Namiki,T., Ohmeda,E., Yahagi,W., Suzuki,K., Li,C.,						
		Kojima,K., Shishiki,T., Foundation of Advancement of International						
		Science Genome Sequencing & Analysis Group; Otomo,Y., Murakami,K.,						
		Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,						
		Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,						
		Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niihara,J.,						
		Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,						
		Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN;						
		Kawai,J., Carinici,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,						
		Hare,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,						
		Kagewa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,						
		Saito,R., Saeki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,						
		Yoshino,M., and Hayashizaki,Y.						
TITLE		Collection, mapping, and annotation of over 28,000 cDNA clones from						

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

japonica rice
Science 301 (5631), 376-379 (2003)
22752273
12869764
2 (bases 1 to 3455)

ADACHI, J., AIZAWA, K., AKIMURA, T., ARAKAWA, T., CARNINCI, P., DOI, K.,
FUJIMURA, T., FUKUDA, S., HANAGAKI, T., HARA, A., HASHIZUME, W.,
HAYASHIDA, K., HAYASHIRAKI, Y., HAYATSU, N., HIRAMOTO, K., HIRAOKA, T.,
HORI, F., HOTTA, I., IIDA, J., IIDA, Y., IKEDA, R., IMAMURA, K., IMOTANI, K.,
ISHIBIKI, J., ISHII, Y., ISHIKAWA, W., ITOH, M., KAGAWA, I.,
KANEGAWA, S., KATO, H., KAWAGASHIRA, N., KAWAI, J., KAWAMATA, K.,
KIKUCHI, S., KISHIKAWA-HIROZANE, T., KISHIMOTO, N., KOBAYASHI, M.,
KOMADA, T., KOJIMA, K., KOJIMA, Y., KONDO, S., KONNO, H., KOUDE, M.,
KOYA, S., KURIHARA, C., KUROSAKI, T., KUSUMEGI, T., LI, C., LU, M.,
MAEDA, H., MATSUBARA, K., MATSUYAMA, T., MIURA, J., MIYAZAKI, A.,
MIZUNO, K., MURAKAMI, K., MURAKA, T., NAGATA, T., NAKAMURA, K.,
NAMIKI, T., NARIKAWA, R., NIKIKURA, J., NISHI, K., NOMURA, K.,
NUMASAKI, R., ONEDA, E., OHNO, M., OHTSUKI, K., OKA, M., OOKA, H.,
OSATO, N., OGA, Y., OTOMO, Y., RYU, R., SATOH, H., SAKAI, C., SAKAI, K.,
SAKAZUME, N., SANO, H., SAKAKI, D., SATO, K., SATOH, K., SHIBARA, K.,
SHINAGAWA, A., SHIRAKI, T., SHISHIKI, T., SOGABE, Y., SUGANO, S.,
SUGIYAMA, A., SUZUKI, K., SUZUKI, Y., TAGAMI, M., TAGAMI-TAKEDA, Y.,
TAGAWA, A., TAKAHASHI, F., TAKAKU-AKISHIRA, S., TANAKA, T., TOMARU, A.,
TOYA, T., TSUNODA, Y., UEDA, M., WAKI, K., XIE, Q., YAHAGI, W.,
YAMADA, H., YAMAMOTO, M., YASUNISHI, A., YAZAKI, J., YOKOMIZO, S. and
YOSHIDA, A.

Direct Submission
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression(2-1-2 Kamondoji, Tsukuba, Ibaraki
305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007)
This clone is one of the 28K full-length cDNA clones from japonica
rice.

TITLE
JOURNAL
COMMENT

URL : http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-length_cDNA_Project Gene Project Name: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Oneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Otomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Nikikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken
ADACHI, J., AIZAWA, K., AKIMURA, T., ARAKAWA, T., CARNINCI, P., FUKUDA, S., HANAGAKI, T., HARA, A., HASHIZUME, W., HAYASHIDA, K., HAYATSU, N., HIRAMOTO, K., HIRAOKA, T., HORI, F., IIDA, J., IIDA, Y., IMAMURA, K., IMOTANI, K., ISHII, Y., ITOH, M., KAGAWA, I., KANEGAWA, S., KATO, H., KAWAI, J., KISHIKAWA-HIROZANE, T., KOJIMA, Y., KONDO, S., KONNO, H., KOUDE, M., KOYA, S., KURIHARA, C., MATSUYAMA, T., MIYAZAKI, A., MURATA, M., NAKAMURA, M., NISHI, K., NOMURA, K., NUMASAKI, R., OHNO, M., OSATO, N., OGA, Y., SATOH, H., SAKAI, C., SAKAI, K., SAKAZUME, N., SANO, H., SASEAKI, D., SATO, K., SHIBATA, K., SHINAGAWA, A., SHIRAKI, T., SOGABE, Y., TAGAMI, M., TAGAMI-TAKEDA, Y., TAGAWA, A., TAKAHASHI, F., TAKAKU-AKISHIRA, S., TANAKA, T., TOMARU, A., TOYA, T., WAKI, K., YASUNISHI, A. and Hayashiraki, Y.

Location/Qualifiers

FEATURES
source
1. 3455
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultiyar="Nijponbare"
/db_xref="taxon:39947"
/clone="U023129A05"

ORIGIN

Alignment Scores:
Fred. No.: 2.05e-23 3455
Score: 799.00 Matches: 298
Percent Similarity: 37.20% Conservative: 119

Best Local Similarity:	26.58%	Mismatches:	366
Query Match:	16.64%	Indels:	378
DB:	8	Gaps:	45

QY	57	ProPheArgLAlaAspAspSerGlyPhe-----GlnSerAsnAsnLeuTrpVal	22
Db	203	CCGCGCGGCGGCTTCCTCCCGGGGGGGGATGGACCGCGCTGGCGACCTGGGGTG	263
QY	23	GlySerLeuThrProGluTrhTrhGlnSerAspLeuThrGluLeuPheGlyArgGly	42
Db	263	GGCAGCGCTCCCGCGGGCGTGC CGCGCGGCGGACCTTCGGAGCTCTTC CGGGTGCGC	322
QY	43	AspLLeaAspArgLLeThrValTy rSerSerArgGlyPheLLePheLLeTy rTy rArgHis	62
Db	323	GAGGTGGAGGCACTTCCTCCGTAA CCGCGCGGAGCTTCGCTTCGAGCTTCGCGCGG	382
QY	63	ValGluGluLalaValAlaAlaTy rGluLLeuGlnGlyLalaAsnLeuAsnGlySerGln	82
Db	383	GAGGAGGACGCGCTGGCGGCGGCTGGCGGAGCTGCAGGCGGATCCACTCCGCGGGCGCCC	442
QY	83	LLeTy rLLeGluTy rLalaArgProAlaTy rProCysLeuSerLeuTrpValGlyGlyLe	102
Db	443	ATTGAGATGAGATTTCACAGGGGAGATTAAGATTCAAGTAGC-----	484
QY	103	GlyProAlaValSerTy rAspAspLeuGluGluGluPheSerTy rPheGlyLysLLeGlu	122
Db	485	-----TCATGGATGAC-----AGATTACTCAACAATGCT	514
QY	123	AspPheArgPheLeuArgGluArgTy rLTrhAlaPheLLeAspTy rTy rGluLeuMetAsp	142
Db	515	GATCAAGACGTTTACTGACACAGGA-----	541
QY	143	AlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAsp	162
Db	541	-----	541
QY	163	PheLeuArgSerGlnAlaProLysTy rGluGlnTrpAlaGlySerTy rAspAsnAlaAsn	182
Db	542	-----AGGAATTCAGCATCAAGTCTGTAAAA-----TCAACTGAT	577
QY	183	GlyAsnMetAsnHisLysSerProGlnTy rProHisSerTy rGluAspPheLysGlyAspVal	202
Db	578	AAATCCAAAAGAGCAGGCCA-----GCA	601
QY	203	GlnProSerLysValLeuTrpLLeGlyPheProProThrAlaThrGlnCysAsnAspGln	222
Db	602	GAACTAGTGAAGTATATGATGATGGTTTTCTCGTGTGCTGAAGTA-----GATGAG	655
QY	223	GlnLLeLeuHisAsnAlaMetLLeLeuPheGlyGlnLLeGluArgTy rLysSerTy rPro	242
Db	656	GCAACTCTCTGGGAAGCCTTTTCACTTTTGGTGAAGGTGTCAAGTAACTACATTTCCCA	715
QY	243	SerAlaArgAsnPheAlaLeuValGluPheArgSerAlaGluGlnAlaArgGlnCysLysGln	262
Db	716	GGCGCTACTTATTCATTTGTCCAGTACACTACTATTTCGACGCGCATGCGCGCGAAGGAA	775
QY	263	GlyLeuGlnGlyArgLeuPheAsnAsnProArgLLeTy rLLeMetTy rSer-----	279
Db	776	ACACTGCAGGGAATAATTTTCAATAACCTCGAGTTAGCATTTGCTTTCTCGAGTGC	835
QY	280	-----Asn	280
Db	836	AGTGTTCAGAGAAATTGGAAAAGTTCCTTAGATGCCCATATTCCTCCCATTTTAAAC	895
QY	281	AspGluLeuProPro-----GlnGlnAspAspTrhSerPheTy rSerLysLysLys	297
Db	896	TCTAGTGTAGACCTATATATTCAGGAGCAAGTTTGTGAAGATTT-----	940
QY	298	ArgSerAlaGlyTrhAspMetPheAsnAsnAspPro-----SerCysValSerSerProHis	315
Db	941	---CCTAAGGCTAAGCCTTTGATGATGCTCTCCAGAGATATGTCAATCATTCATTCACAT	997

QY	316	SeThrGlyIleProGlySerMetArgProLeuAlaGlyIleThrAsnGluArgSerTyrAsn	335
Db	998	TAT-----GGCCCTAAGAGACTTCTTGA-----	1022
QY	336	GIAlaGluTyrAsnAspValValGlyGluGluProAsnTyrArgArgProSerAlaAsn	355
Db	1022	-----GATCATGATGATATGGGTTTCAGACAGGATAAATTATTGGGA-----	1066
QY	356	GlyThrGlyIleLeuProSerProThrGlyIleLeuProSerPro-----	372
Db	1064	-----TATGACCTGGAGTAGAGCCCTGATCTGATCTAAT	1099
QY	373	-----AlaGlnGlyThrArgProMetArgSerAsnProAspSerTyr	387
Db	1100	TTTGAACCTTTTAGATACAGAGGCTCGGTCCAGAAAGAGATGCTGAGAGCCCATAT	1155
QY	388	GluGlyTyr-----	390
Db	1160	GAACAGCATAGCGCTAGCCCTGCTGGTATGACACATGGCACACATTCATTCAGGGA	1211
QY	391	-----AspProAlaGlnLeuValArgGlu-----	398
Db	1220	TCTCAGGAGCCTTACCATTAAGAGATTCCTCGTATGCTAGGGAAGATCATACCATTT	1279
QY	399	SerTyrArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGlu	418
Db	1280	TCAAGAAGATTGAGACACTGTGTAACACATGACTGTGAATTCTCT--GAATACCTTTTC	1338
QY	419	ArgSerPheGlyArgGlySerValAlaAla-----ArgProIleArgGlyPro	434
Db	1337	TCTGAATTTGATCGAGGAGAGTTGCTCTGCTCTACCCAGAGAGCCCTTCTATGGTGTG	1398
QY	435	ProAspSerAsp-----	438
Db	1397	CCAGTATGATGACATACACCCAGAGGCTATCAACTTGTCTTATGATGATGAGAAATCAT	1456
QY	439	-----HisIle-----	440
Db	1457	GTTGATCCTTAAAGAAATCCAATCTCATTTAGATAGACATATACAGGCGATGACAG	1516
QY	440	-----	440
Db	1517	GACAGCTTTTCTAGCATGTAGAAAGTGAAGAAATCAACTCTGAATACATGAACCCCTT	1576
QY	441	-----TyrArgGlyMetIleAlaGlyGlyIleThrProValCysCys	454
Db	1577	CTCAAGAAAGATGGAATGGGATGGATGATACATAGCAAAAGGAGGACACCAATTGGCGA	1636
QY	455	AlaArgCysValProMetGlyIleGluThrIleLeuProGluValAlaAsnCys	474
Db	1637	GCGCATCTCTCCCTGTTGGGAGAAAGTCTTAACTTCATGCTGCCCGAATTTTGGATTGC	1696
QY	475	SerAlaArgThrAspLeuAsnMetLeuAlaIleHisTyrAlaValAlaIleGlyCysGlu	494
Db	1697	ACTGTAGAGCAAGCTGTGAGATGCTCTTAGACATTAATTCAGAGCTCCAGACGCTGG	1756
QY	495	IleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArg	514
Db	1757	GTGGTGTTTTGTTCAGAAAAGATGTGACATGCGACAGCCCTATATGATTGATCATAT	1816
QY	515	TyrIleuSerSerIleAspArgAlaGlyValAlaIleLeuAsnAspGlyThrThrLeuPhe	534
Db	1817	TACCTTGTGTAAAGACGCTGACAGAGTTGTAACTTGAGAAAGACAGACGTTATTT	1876
QY	535	LeuValProProSerAspPheLeuThrAspValLeuGlnIleThrArgGlnIleArgLeu	554
Db	1877	CTGTGTACACCTCAGACTTCTTGAACAAGTACTAGAGGTTCCAGGTAAGTCAGACATA	1936
QY	555	TyrGlyValValLeuIleuPheLeuProProProAlaValProValThrAlaSerTyrArgGln	574
Db	1937	TCTGAGTCATCTGAAGTTTGACAGCTGACATCCAGAAATTTCTCTGCCAATCTGGCAA	1996

QY 575 Gluser----- 576
 Db 1997 CCGAAGACATTGTGAGTCATTGAAACCATGATGTCCTGCTATGAGATCTTAAGATCA 2056
 QY 577 -----GlnSerAsnPro-----LeuHisTyrMetAspGlnAlaArgAsp----- 589
 Db 2057 TTGAGAAAGATCAACCCACAGATATCAGGCCACTTCCTCAGGGTTTCAAGTTATTCGGG 2116
 QY 590 ---SerProAlaAsn-----AlaSerHisSerLeuTyrProProAlaGluAsnTyr 605
 Db 2117 TTGTGGCTGGAGCTGATTAATCCAGCAAGTGCACATTGGTTCGCTTACAG----- 2170
 QY 606 IleArgGlyAlaProGlnHisLeuThrAla----- 615
 Db 2171 TTGGAAATGCTCTTATATCTGAATCTGAATTAGCTCATGAAGATCCACCTGAC 2230
 QY 616 -----AlaSerTyrProSerValSerGluProLeuArgIleProAsn--- 629
 Db 2231 TCCACAGAGGAGATGACATGACAGCAGCAGCAACCCAGATGTATGGCTCAGAGA 2290
 QY 630 -----AsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeu 643
 Db 2291 TGGTCAGATTAACATTACATCCAGTCCAGGTTCTGGA-----AATTG 2335
 QY 644 LeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHis 663
 Db 2336 AATTATTTGGCTGAGAGTGGATGCCATCCACATACATCACTGATAGACACAGAGGCGATAC 2395
 QY 664 Gln-----PrometSerGlyProSerThrValValSerThrAlaHisGlnSerAsnGly 681
 Db 2396 TCATTGCTCTCCTCAAGAGATACCAAA-----GTGAGTACA----- 2431
 QY 682 LeuTyrAsnGlyGluAlaProSerGlnAlaTPrLysArgGlyProGlnThrValHisAsp 701
 Db 2432 -----TCAGGGTATGACCA-----GTTGACAGATGAG 2458
 QY 702 AlaSerAsnGlnSerPhe-----GlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGln 719
 Db 2459 GCATCAACACATGCTCTACCTCCACATGCCAATCTGCATACAGCAGGTATTAACCTCA 2518
 QY 720 LeuProProPro---ProSerArgTyrProPro----- 729
 Db 2519 CAACCTTCATCTCTCCCATATATGGCTTCCACAGAGCAATTGCACAATTGGCCACTCTT 2578
 QY 730 -----AlaSerAsnAsnPro-----AsnTyrThrSerGly 739
 Db 2579 CTTCGACAAACAAACCAACCAAGCAAAAGACCTGTGACAGCTTAAACAAAGATCAGGA 2638
 QY 740 MetVal-----HisGlyAsn-----MetGlnTyrGlnSerGlnSerValAsn 753
 Db 2639 TTCATACGGAATCCACATGACATCTTCTTAATATGATGCCACAGCCTCAGGTTCTATC-- 2695
 QY 754 MetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGln 773
 Db 2696 ---CTGTCTCAAAACTATTCATTCGCCACCTGCTCG-----CCA 2728
 QY 774 GlySerSerAsnHisProValSerGlnProMetValGln----- 786
 Db 2729 TCTGCATCAGAGTTACAGTTTACATGCGCCACAGTTCAGTTCAGTTCAGCAAAATCCA 2788
 QY 787 -----GlnTyrGlnProGluAlaSerMetProAsnGlnAsnTyrGlyProIleProSer 804
 Db 2789 TCCATTATGCAATCAGCAAGATGCTCTTATGCTTTCACAAACCTTTCAGTTCAGTTCAGT 2848
 QY 805 TyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnLeuLeuAsnProSer 824
 Db 2849 ATGATCTCTCGGGA-----AATCAGCTCATCTTCATCAGCTCTTG 2890
 QY 825 GlnPheGlnAlaAlaMetGlnProProAlaAspLysLeuAsnLeuGluProGlnAsnGln 844
 Db 2891 AGATCGTTCTGCTCTCTCTTCTGAAAGGCTTCACCCCTTAGGCGAGCACATCAAGT 2950
 QY 845 AlaLeuArgLeuGlnProMetIleSerGlyAspGlyGln-----Gly 858

Db 2951 GCTTACAGCCCACTCCTCCTCTGAGACCAAGATCAAGTACCAAGTCTGCT 3010
 QY 859 ThrThrAspGlyGluValaAspLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAla 878
 Db 3011 CAGAAAGACCATTCACAGAGACCTCCAAAGCCCTTCAAGCAATTCGAATTGGCAGCA 3070
 QY 879 AsnLeuLeuGlnGlnIleGlnGlnTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 898
 Db 3071 ACCCTACTTCACAGATACAGAACATCTAA-----CCTGCTGC 3112
 QY 899 Gln 899
 Db 3113 CAG 3115
 RESULT 13
 AX420424 3715 bp DNA linear PAT 18-JUN-2002
 LOCUS AX420424
 DEFINITION Sequence 6 from Patent WO0212518.
 ACCESSION AX420424
 VERSION AX420424.1 GI:21524580
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1
 REFERENCE
 AUTHORS Amesino, R.M., Schomburg, F.M., Michaels, S.D. and Patton, D.
 TITLES Floral induction gene
 JOURNAL Patent: WO 0212518-A 6 14-FEB-2002;
 WISCONSIN ALUMNI RESEARCH FOUNDATION (US)
 FEATURES
 source
 1..3715
 /organism="Arabidopsis thaliana"
 /mol_type="unassigned DNA"
 /db_xref="taxon:3702"
 /note="PPA promoter plus intron"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,786-17 Length: 3715
 Score: 651.50 Matches: 171
 Percent Similarity: 27.238 Conservative: 0
 Best Local Similarity: 27.238 Mismatches: 0
 Query Match: 13.574 Indels: 457
 DB: 6 Gaps: 1
 US-09-920-705-3 (1-901) x AX420424 (1-3715)
 QY 1 MetAlaLeuSerMetIleProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
 Db 1832 ATGGCGTATTCATGAGACCATTCAGAGCCGATGATTCGGTTCACAGCAAAATCTT 1891
 QY 21 TrpValGlySerLeuThrProGluThrThrArgIleAspLeuThrGluLeuPheGlyArg 40
 Db 1892 TGGGTGCTGATGCTTAACGCGGAGAGCAGACGATCAAGTCAAGTGTGTTGGAAAGA 1951
 QY 41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
 Db 1952 TACGGGATATTCATGATGATCACGGGTATTTCTTACAGAGGCTTTCGTTATATATCTAC 2011
 QY 61 ArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
 Db 2012 AGACATGTGAGAGAAAGCAGTCCAGCCAAAGAGGCTTTCAGAGGCAAAATTTGAATGA 2071
 QY 81 SerGlnIleLysIleGluThrAlaArgPro----- 90
 Db 2072 AGTCAATTAAGATGAATACGACGACGCGTTGTTCTTATCTATCTTCGCTTGTTC 2131
 QY 90 ----- 90
 Db 2132 TCTAATTGATGTCTTTTGTCAACGATTAATCTTTTTCGAATTCATGTCAGGT 2191

QY 90 ----- 90
 Db 2192 TCACAAACTTGTATGATGCTTGTAGTCCAAAAATTTCTTGTGATCTGTTTTTTT 2251
 QY 90 ----- 90
 Db 2252 TTCTCATCGTACAAATCAAGTCGAAACCTAGTTTTTTCTATTATAGTCGTAGCTT 2311
 QY 90 ----- 90
 Db 2312 AAGCGAACTGTATCCGATCGAAAGCTCTTTCTCAATTACTTTGGTTAATCGAACT 2371
 QY 90 ----- 90
 Db 2372 CGCGCAAGCCAAACACAGAGAGCTCTGAAAATTTGATGTTAAAGCATATATACTC 2431
 QY 90 ----- 90
 Db 2432 TTAGCGAATGAGCTCTGCAAAAGATCATCAACATGTTTACTCTCGTTATGCGAA 2491
 QY 90 ----- 90
 Db 2492 GAAGTTAATCCGATTCGTTTATCTGCAAACTTGATCCGCTTTAGCACTTGANT 2551
 QY 90 ----- 90
 Db 2552 TCAGTTCGCTGATGTTAAATCTCAAGCTTGAGTATGATGATTCGCTACTCCA 2611
 QY 90 ----- 90
 Db 2612 CTGCAATTCGCAATGCTGAAAAATTGGAGCGAGCAAACTCTATCATCGGCCAAACAG 2671
 QY 90 ----- 90
 Db 2672 TAAGAAATTTTGAAGTTTAAAGTTCAGTTTCTCGCAAAAAATCAACCGGTGGAGAAAT 2731
 QY 90 ----- 90
 Db 2732 TTGTCTATGCGGTAGCTAGATATCAATACTGCATCTGGCAACAGAAAAATTTGCGAA 2791
 QY 90 ----- 90
 Db 2792 CCGTATGCTCTTAATCTCTAGTATGCTGCGACATATCATCATGCTGAGGAGTTT 2851
 QY 90 ----- 90
 Db 2852 GGGCTTGTGCTCGATGCTTCAAGAAAGCAAAATTAGTGGTTGTAAGCGGCTAAC 2911
 QY 90 ----- 90
 Db 2912 TCTACCAATCAGAAAGCTTATTTCAGAAACATGTTTCTTCATTTCCATCCCTA 2971
 QY 90 ----- 90
 Db 2972 CGTACATCTGGGCTTCCCATTTGATGTTCTTTAGGAAGTTGACTTTCTTCAACCAATTC 3031
 QY 90 ----- 90
 Db 3032 TGGGATTGTTACAGTTGAGTGAGCAAGCAAAATATTTCATTGACGGAGCACTCAATTT 3091
 QY 90 ----- 90
 Db 3092 CAATGCTTGGCGGCTTTGAGTACTTAAATGAGTGTTCAGGAGTTGTTATTGCGGAC 3151
 QY 90 ----- 90
 Db 3152 TCAGCAATGATTAAGACAGAAATGTTTGCAGAAAGTATTAAATGTTCTTTCGCGCTTG 3211
 QY 90 ----- 90
 Db 3212 GGGACTACATCTTCTCTACAGACAAAAATCTGATGTTTAGGACAGAGAACTATTACA 3271

QY 90 ----- 90
 Db 3272 CTGATTTAATAGAGAAAGAGAGATGTTTCTCTGGCGCAATTTATTGTTTAGAA 3331
 QY 90 ----- 90
 Db 3332 AGCAATTGATAGAAATGSGTGTCTAGTATGTTGAATTAATTAAGTTTGTCTGTT 3391
 QY 90 ----- 90
 Db 3392 AGTTCTCTGATGTTTGTATGCTTATCTTGGCACTATCTGGATAGTATGCTTCTT 3451
 QY 91 ----- 91
 Db 3452 ATCGACTTTTCTTGTGACGCAAACTTGTAAGATCTATGAGTGGTGGAAATCGGC 3511
 QY 104 ProAnaValSerLyAspAspLeuGluGluPheSerLyPheGlyLeu11GluAsp 123
 Db 3512 CTAATGCTCCAGAGATGACCTGGAGAGAGGTTCAACAGTTTGGGAAATCGAGAT 3571
 QY 124 PheArgPheLeuArgGluArgLyThrAlaPhe11LeuAsp1YrYrGluMetAspAspAla 143
 Db 3572 TTTAGGTTTCTCAGAGAACGAGACAGCTTTCATTTATTTATGATGATGATGATGCT 3631
 QY 144 LeuGlnAlaLySerMetAspGlyLeuProMetGlyGlySerPheLeuArgValAspPhe 163
 Db 3632 TTACAGGCTTAAGAGATGAATGAAAGCTTATGGGTGATGCTTTTGGCTGTGATTTT 3691
 QY 164 LeuArgSerGlnAlaProLyAla 171
 Db 3692 CTCCGCTCAACAGCGCCAAAAAAA 3715

RESULT 14
 AP004011/c
 LOCUS
 DEFINITION
 Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 9,
 BAC clone:OUJ381_H04.
 AP004011
 ACCESSION
 VERSION
 KEYWORDS
 ORGANISM
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Superkaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.
 REFERENCE
 1
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, BAC
 clone:OUJ381_H04
 Published Only in Database (2001)
 2 (bases 1 to 130961)
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Direct Submission
 Submitted (01-AUG-2001) Takuji Sasaki, National Institute of
 Agrobiological Sciences, Rice Genome Research Program; Kamondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 On Jul 27, 2004 this sequence version replaced gi:46092388.
 Genes were predicted from the integrated results of the following:
 GENSCAN (http://ccr-081.mit.edu/GENSCAN.html), FGENESH
 (http://www.softberry.com/), GeneMark.hmm
 (http://opal.biology.gatech.edu/GeneMark/), GLIMMER
 (http://www.tigr.org/cdb/glimmer/glimr_form.html), RICEHMM
 (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
 (http://biointformatics.iastate.edu/cgi-bin/sp.cgi), slm4
 (http://globin.cse.psu.edu/html/docs/slm4.html), gap2
 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The
 genomic sequence was searched against NCBI Nonredundant Protein
 database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
 sequence database at RGP or DDBJ. Protein homologues of the coding
 regions were searched against NCBI Nonredundant Protein database
 with BLASTP. ESTs represent the identified cDNA sequences using

BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to INGS standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence. The orientation of the sequence is from M13rev to -21M13 of the BAC clone. This sequence of OJ1381_H04 clone has an overlap with F0592C05 (DDBJ: AF004756) clone at 5' end and an overlap with OSUNBA0062409 (DDBJ: AF004736) clone at 3' end. The sequence was generated by combining Monsanto and RGP-Japan sequencing data. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rtp.dna.affrc.go.jp/GenomeSeq.html>.

FEATURES

source
1. .130961
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="9"
/clone="OJ1381_H04"
/complement(3074..6118)
/gene="OJ1381_H04.1"
/complement(join(3074..3542,3637..3762,4173..4312,5566..6063))
/gene="OJ1381_H04.1"
/note="supported by full-length cDNA(s) : AK104965"
/complement(join(3105..3542,3637..3762,4173..4312,5566..6118))
/gene="OJ1381_H04.1"
/note="supported by full-length cDNA(s) : AK073729"
complement(join(3402..3542,3637..3762,4173..4312,5566..6005))
/gene="OJ1381_H04.1"
/note="contains EST(s) : AU174422(S6559), C25488(S6559) contains full-length cDNA(s) : AK104965, AK073729"
/codon_start=1
/product="putative ubiquitin-conjugating enzyme family protein"
/protein_id="BAD32975.1"
/db_xref="GI:50725506"
/translation="MTLGSSGAGSSVVRNRLLELRGKGTGDTVSQMDAD DTKRSWTITIGPHNSVHEGRVQLKCDKDYEPKPSVRPHSRINNSCVNHETGA VSRKRGVLANQREYTMETITLTQLKKNATPDQNRLLVQPPGTFP"
/complement(8645..9266)
/gene="OJ1381_H04.2"
/complement(8645..9266)
/gene="OJ1381_H04.2"
/note="probably inactive due to 3'exon missing in CDS pseudogene, serine/threonine phosphatase PP7-related"
/pseudo
complement(9412..10109)
/gene="OJ1381_H04.3"
/complement(9412..10109)
/gene="OJ1381_H04.3"
/note="probably inactive due to 3'exon missing in CDS pseudogene, transposase related protein"
/pseudo
complement(14903..15292)
/gene="OJ1381_H04.4"
/complement(<14903..>15292)
/gene="OJ1381_H04.4"
/note="start and end point are not identified"

CDS
complement(14903..15292)
/gene="OJ1381_H04.4"
/note="similar to Oryza sativa chromosome 7, P0669H03.22"
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD32976.1"
/db_xref="GI:50725507"
/translation="MASRWVATARTKSGRCDCRGORWRRLSPPTTMASTRIAAAC SGRQWWRERKRRKRWEGEDTDT"
join(16127..16171,16261..16372,16416..16460,16533..16753)
join(16127..16171,16261..16372,16416..16460,16533..16753)
/gene="OJ1381_H04.5"
/complement(16127..16171,16261..16372,16416..16460,16533..16753)
/note="hypothetical ORF predicted by GENSCAN this category is not included in INGS standard"
16879..18432
/gene="OJ1381_H04.6"
/gene="OJ1381_H04.6"
/note="start and end point are not identified"
16879..18432
/gene="OJ1381_H04.6"
/codon_start=1
/product="putative monosaccharide transporter 6"
/protein_id="BAD32977.1"
/db_xref="GI:50725508"
/translation="MVVASGAGRPYPGGLTFVSMACLVAAATGGLIFGYDIGSGV TSMDFLSRFPSVYRAOASAAAGAGQYCFPDSQLTFPSSLYLAALASLGAT VTRVAGRKMSMPAGGLVFLAGCALGAAVAVMLVGRVLLGVGFANQSVVYVTS MA PARMGILNNGPOMMTTGYLANLNYCARFAGCGWRLSLAALVPAVATAG ALFIDETNSLIERRGRBARMLORVGEYDMDENDVLAAGASIVASPRDI LTRNRPLVAVAIPLFOQLGVINVEYAVLRTLPFGSGLMSAVITGGVMA ATLVAVLAVRVGRALPLRGAQVAVAOAALIGRLFGSGLMAIPAGVAAVVA MACVVAAPFAMSGPLAMLVSEVWFLVRPAGQSTVAVNNAATFAVAQATPLICR LRFVFPFPAAGVAAATFVALFVETKGVPIEDMAAVMSDHMYKRFVDDGDGAR RGDIMGHK"
complement(20105..23575)
/gene="OJ1381_H04.7"
/complement(join(20105..20513,20597..20726,20898..20966,21052..21171,21461..21545,21670..21755,21867..22019,22113..22235,23079..23575))
/gene="OJ1381_H04.7"
/note="supported by full-length cDNA(s) : AK071036"
complement(join(20437..20513,20597..20726,20898..20966,21052..21171,21461..21545,21670..21755,21867..22019,22113..22235,23079..23462))
/gene="OJ1381_H04.7"
/note="contains EST(s) : AU100795(C51066), C27141(C51066) contains full-length cDNA(s) : AK071036"
/codon_start=1
/product="phosphoenolpyruvate/phosphate translocator"
/protein_id="BAD32978.1"
/db_xref="GI:50725509"
/translation="MQSAAYGLRPPCATTAAPLQLRNPSRPGVCGQPLAPR GLRLSAVPRAGTSKRITGLVPASBPQEDERRRGARDAVAATAAGABGEBGGGL AKTIQLGALFGIMVYFNTYFNTYKQVAKVPSYINTVQAVGTVALFEMWITGIL KRPKISGAQALAIPLAVHTGNLFTMNSLQKAVSFTHITKAPSPSVLSALFL GEMPFPVVASLVPVIGVALASLTAEASFNAGVSAASNTVFGSRVLSKGLAWK BESILNITLFSITITVMSFPLAPVTLTREGVCTVPTVQSAGLNKKOYTRSLIAFC PHAYQOVSYMLARVSPYTHSVGNCKVAVVIVSVLFRPTVPSINSLGTVALAGV FLYSQLRKLKPEKTA"
complement(24740..25376)
/gene="OJ1381_H04.8"
complement(join(<24740..24897,25172..>25376))
/gene="OJ1381_H04.8"
/note="start and end point are not identified"
complement(join(24740..24897,25172..25376))
/gene="OJ1381_H04.8"
/note="predicted by Genemark.hmm etc."
/codon_start=1
/product="hypothetical protein"

/protein_id="BAD32979.1"
 /db_xref="GI:50725510"
 /translation="MEVSRGSPDNPSPICQENATSLRRQDAREEREEDLPPEPCGHC
 TREMLRGKARKATTPRRRLAEVTTATLIDLNLCSGGVGSBGAVGATMBEEMV
 EEVVGIGDKVIMLRVSFS"
 complement(26622..27293)
 /gene="OJ1381_H04.9"
 complement(join(26622..26706,26948..27094,27220..>27293))
 /gene="OJ1381_H04.9"
 /note="start and end point are not identified"
 complement(join(26622..26706,26948..27094,27220..27293))
 /gene="OJ1381_H04.9"
 /note="Predicted by GeneMark.hmm etc."
 /codon_start=1
 /product="hypothetical protein"

CDS

Alignment Scores:
 Pred. No.: 8.31e-13 Length: 130961
 Score: 576.00 Matches: 257
 Percent Similarity: 31.28% Conservative: 99
 Best Local Similarity: 22.58% Mismatches: 312
 Query Match: 12.00% Indels: 470
 DB: 8 Gaps: 41

US-09-920-705-3 (1-901) x AP004011 (1-130961)

QY 144 LeuGlnAlaIysSerMetAenGlyLysProMetGlyLysPheLeuArgValAspPhe 163
 Db 95385 CTGGACCTGAATATTTGAATTATTCAGATAAGGTTCAAGTCTATGATGCACAA 95326
 QY 164 LeuArgSerGlnAlaProLysLysGluGlnTTPAlaGlySerTyrAspAsnArgAenGly 183
 Db 95325 TACTCACACATCTGTATCAAGACGT-----TTTACTGAACGAGGAGAG 95281
 QY 184 AsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspValGln 203
 Db 95280 AATCAGAGATCAAGTCTGTAATAATCACTGATTAATCCAAAAGACGCCACAGCA 95221
 QY 204 ProSerLysValLeuTTPLeuTTPLeuTTPLeuTTPLeuTTPLeuTTPLeuTTPLeu 223
 Db 95220 CCTAAGAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 95167
 QY 224 IleLeuHisAsnAlaMetIleLeuPheGlyIleGluArgValLysSerTyrProSer 243
 Db 95166 ACTCTCTGGGAACCTTTTTCACCTTTGGTGGAGTTGTCAAGATTAATCAATCCACAG 95107
 QY 244 ArgAsnPheAlaLeuValGluPheArgSerAlaGluAlaArgGlnCysLysGly 263
 Db 95106 CGAATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 95047
 QY 264 LeuGlnGlyArgLeuPheAsnProAlaGlyIleLysIleMetTyrSer 279
 Db 95046 CTGCAGGGAATATTTTCAATTAACCTCGATTGATGATGATGATGATGATGATGAT 94987
 QY 280 -----AsnAsp 281
 Db 94986 GTTTCAGCAGATTTGAAAAGTTCCTTAGTCCCATATTTCCCATTTAAACCT 94927
 QY 282 GluLeuProPro-----GluGlnAspAspThrSerPheTyrSerGlyMetLysArg 298
 Db 94926 AGTGTAGACCTATATTCAGAGCAGCAAGATTTTGAAGATTT----- 94885
 QY 299 SerArgThrAspMetPheAsnAsnAspPro-----SerCysValSerSerProHisSer 316
 Db 94884 CCTAAGGCGTACGCTTTTGAATGATGCTCCAGAGATATGATACATTCCTCCACATTA 94825
 QY 317 ThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGly 336
 Db 94824 -----GGCCCTTAAGACCTTCTAGA----- 94804
 QY 337 AlaGluTyrAsnAspValValGlyLysGluProAsnTTPArgArgProSerAlaAsnGly 356
 Db 94803 ---GATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 94762

QY 357 ThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerPro----- 372
 Db 94761 -----TATGACCTGAGATAGAGCCCTGATCTTATGATCTTAATTT 94723
 QY 373 -----AlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTTPGlu 388
 Db 94722 GAACCTTTTAGATATCAAGGGCTCGGTCCAGAAAGAGATGTCTGAGACCCATATGAA 94663
 QY 389 GlyTyr----- 390
 Db 94662 CAGCATAGCCGTAAGCCCTGCTGTATGATGACCATGCAACATTTCCATTGACGATCT 94603
 QY 391 -----AspProAlaGlnLeuValArgGlu-----Ser 399
 Db 94602 CAGGAGCCTTACCATTAAGGATTTCTCGTATGCTAGGAAAGATCCATACCATTTTCA 94543
 QY 400 LysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArg 419
 Db 94542 AAGAAAGTTGAGGACTGTGAAAGCACATGACTGAACTTCTCT--GAATACCTTTCTCT 94486
 QY 420 SerPheGlyArgGlySerValAlaAla-----ArgProIleArgGlyProPro 435
 Db 94485 GAATTTGATCGAGGAGAGATTTGGCTCTGCTCCATACCCAGAGAGCCCTTCTATGCTGTGCA 94426
 QY 436 AspSerAsp----- 438
 Db 94425 GATGATGACATACACCCAGAGCTATCACTGCTCTATGATGATGATGATGATGATGAT 94366
 QY 439 -----HisIle----- 440
 Db 94365 GATCCTTTAAGGAATCCAACTCCACTGTGATGATGATGATGATGATGATGATGATGAT 94306
 QY 440 ----- 440
 Db 94305 AGCTTTCTAGCATGTAGAAAGTGAAGATCAACTCTGATATACATGAACCCCTTCTC 94246
 QY 441 -----TTPArgGlyMetIleAlaIleGlyLysIleThrProValCysGlyAla 455
 Db 94245 AAGAAAGATGGAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 94186
 QY 456 ArgCysValProMetGlyLysGlyIleGluThrLysLeu----- 468
 Db 94185 CGATGCTTCCCTGTGGGAAGGTTCTTAATCTGATGCTGAAGTCTGGGTTCTTCACTGT 94126
 QY 468 ----- 468
 Db 94125 ATTTCTTGTTCACTTGAATATCATTTATGACATCTACAAATAGCTCTGTATCGT 94066
 QY 468 ----- 468
 Db 94065 CTTAATAAATCACAAAATTTCTCTTCAAGTTCAAGATATATATATACAAATATATA 94006
 QY 468 ----- 468
 Db 94005 TCATAGATATCGTTTGTCTTACATCTAATTTATGATGATGATGATGATGATGATGAT 93946
 QY 468 ----- 468
 Db 93945 ATGAAGCTTTTGAAGCTAGTGTGTTGTTCAAGTTGTTGATTTAATATATCTCTG 93886
 QY 468 ----- 468
 Db 93885 CACCATTTGCTCTACTCATTAATCTTAATAAATACAGAGTTTGCTTCAAGGTCCAAAC 93826
 QY 468 ----- 468
 Db 93825 TACATATCTAGTTGATCAATTTCTATCTGATCTTTTGGTGTATCTGATGATGATGAT 93766
 QY 468 ----- 468
 Db 93765 TTGTGCTTTATGCTCTGTCTTATTTGTTCTTATGATTTGAGCATCTCTAAGTTCAACATTTA 93706

QY 469 -----ProgluValValaenCySerAlaArg 477
 Db 93705 GTTTTAAACGGCTGTTGATTCTTACTTTCAGGCCGAAATTTTGGATTGACCTGCTAGG 93646
 QY 478 ThrAspLeuAsnMetLeuAlaValSHISFYAlaValAlailegLYCySgluileValPhe 497
 Db 93645 ACAAGCCTGGAGATGCTCTCTTAAGCACTATTACCAACCTGCCAGCAGCTGGTGTGTTT 93586
 QY 498 PheValProAspArgIleuGluAspPheAlaSerTYrThrGluPheLeuAlaGTYrLeuSer 517
 Db 93585 TTTGTTCCAGAAATAATGATGCTGATGCGATGATATGATATGATTAATTCAGAAATTCCTTGGT 93526
 QY 518 SerIleAspArgIleValAlaValLeuAspAspGlyThrThrIlePheLeuValPro 537
 Db 93525 GATAGAGAGCTGTCAGAGAGTTTGTAACTTGGAGAAAGAGCAGCTTATTTCTTTGTTCCA 93466
 QY 538 ProSerAspPheLeuThrAspValLeuGlnValThrArgGlnIleValArgLeuTYrGlyVal 557
 Db 93465 CCTCAGACTTCTCTGAAACAAGTACTGAGGGTTCCAGGTAAAGTCAGCATATCTGGAGTC 93406
 QY 558 ValLeuIleuProProProAlaValProValThrAlaSerTYrArgGlnIleuSer 576
 Db 93405 ATTTGAAAGTTTGAGAGTCAGATCCAGAGATTCTCCGCAACTCGCAAAACAGAAACA 93346
 QY 577 -----Gln 577
 Db 93345 TTTGTAGTCATTGGAACCATGATGTTCTGCTCATGAGATCTAGATGATGATGAGAAAGA 93286
 QY 578 SerAsnPro-----LeuHISYrMetAspGlnAlaArgAsp-----SerPro 591
 Db 93285 ATCAACCCAGCATATCAGAGCCACTTCTCAGGGTTTCAGATATATCTCGGGTTGTGCTT 93226
 QY 592 AlaAsn-----AlaSerHisSerLeuTYrProProArgIleuAsnTYrIleArgGly 608
 Db 93225 GGAAGCTATATATCCAGCAAGTGCACATTTGTTCCGCTTACAG-----TTTGAAAT 93172
 QY 609 AlaProGluHisLeuThrAla----- 615
 Db 93171 GCTCCTTCATATCTAGAAATCTGAATTAGTCATCAAAAGATCCACTGCCACAGG 93112
 QY 616 -----AlaSerIleProSerValSerGluProLeuArgIleProAsn----- 629
 Db 93111 GAGATAGCATGACAGCAAGCAGCAACACCAGATGATATGCTCCCAAGATGTCAGAT 93052
 QY 630 -----AsnAlaIleProGlnAlaGlyValSerLeuThrProGluLeuAlaThr 646
 Db 93051 AACATTACATCCAGTCCAGGTTCTGGA-----AATTTGAATTAATTGG 93007
 QY 647 LeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGln----- 664
 Db 93006 GCTGAGAGTCCGATCCCATACATCACTGATAGCAGACACAGAGGATCTCATTTGCT 92947
 QY 665 ProMetSerGlyProSerThrValAlaSerThrAlaHisGlnSerAsnGlyLeuTYrAsn 684
 Db 92946 CCTCAAGAGAGTACCAAAA-----GTGAGTACA-----TCA 92917
 QY 685 GlyGluAlaProSerGlnAlaTyrPlyAsnGlyProGlnThrValHisAspAlaSerAsn 704
 Db 92916 GGGTATGACACA-----GTTCCAGATGAGCATCAAAAC 92884
 QY 705 GlnSerPhe-----GlnGlnTYrGlyAsnGlnTYrThrProAlaGlyGlnLeuProPro 722
 Db 92883 ATGTCTTACCTCCATGCACTGCACTGCACTACAGAGGTAGTACCTCAACAACCTCCA 92824
 QY 723 Pro-----ProSerArgTYrProPro-----Ala 730
 Db 92823 TCTCTCCATTATTCGTTCCACAGAGCACTTGCACAATTTGGCACTCTTTTGACAA 92764
 QY 731 SerAsnAsnPro-----AsnTYrThrSerGlyMetVal--- 741
 Db 92763 CAATACCAACCAAGAAAGAGCTGTGAGCAGCTTAAACAAGAAATCAGATTCATACGG 92704
 QY 742 -----HisGlyAsn-----MetGlnTYrGlnSerGlnSerValAsnMetProGln 756

Db 92703 AATCCACATGAGCATTTCTTAATGATGCCACAGCTCAGGTTATTC-----CCTGTC 92650
 QY 757 LeuSerProLeuProAsnMetProHisAsnAsnTYrSerMetTYrThrGlnIleSer 776
 Db 92649 CAATACCTATTTGCCACCTGCTCG-----CCATCTGATCA 92614
 QY 777 AsnHisProValSerGlnProMetValGln-----Gln 787
 Db 92613 CAGTTACAGGTTTATCGGCACACAGTTCAAGGTTACGTGACGCCAAATCCATTCATTATG 92554
 QY 788 TYrGlnProGluAlaSerMetProAsnGlnAsnTYrGlyProIleProSerTYrGlnGln 807
 Db 92553 CATACACCAAGATGCTCTTATGCTTCTCAACACACTTTCCTTACCCCTTAAGATCTCT 92494
 QY 808 AlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGln 827
 Db 92493 TCGGGA-----AATCCAGCTCATTTCCATGCTTTCAGATGCTTC 92452
 QY 828 AlaAlaMetGlnProProAlaAspIleValAsnLeuGlnProGlnAsnGlnAlaLeuArg 847
 Db 92451 GTCCCTCTCTTCTTAAGGCCCTCCACCCCTTAGGACACACATCAAGTCTTACAA 92392
 QY 848 LeuGlnProMetIleSerGlyAspGlyGln-----GlyThrThrAsp 861
 Db 92391 GCGCAACTGCACTTCCCTTGACACACAGACTAGCCACCAACCCCTGCTCAGAAAGC 92332
 QY 862 GlyGluValAspIleAsnGlnIleArgTYrGlnSerThrLeuGlnPheAlaAsnLeuLeu 881
 Db 92331 CATCAGAGAGACCCCTCAAAAGGCCCTTCAAGCAACATTCATTAATGGCAACAACCTTACT 92272
 QY 882 LeuGlnIleGlnGlnIleAsnGlnGlnIleAsnSerGlyThrProAlaGlyGln 899
 Db 92271 CAGCAGATACAGCAACAATCTAAA-----CTGCTGAGCCAG 92236
 RESULT 15
 ATTPI17
 LOCUS
 Arabidopsis thaliana DNA chromosome 4, BAC clone T1P17 (BSSA
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 137519)
 REFERENCE
 EU Arabidopsis sequencing project.
 AUTHORS
 Direct Submision
 JOURNAL
 Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitze 18a, D-82152 Martinsried, FRG, E-mail:
 lemke@mpi.biochem.mpg.de, mayer@mpi.biochem.mpg.de
 On Sep 5, 1999 this sequence version replaced gi:4725940.
 COMMENT
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.
 Location/Qualifiers
 FEATURES
 source
 1..137519
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /varietal="Columbia"
 /db_xref="taxon:3702"
 /chromosome="4"
 1..18855
 misc_feature
 /note="overlap to BAC T4C9; please refer to this entry for
 analysis and annotation"
 gene
 17536..20463
 /gene="AT4G12420"
 complement(join(17536..17635,17716..17915,18006..18084,
 18165..18530,18612..18978,19061..19199,19285..19555,
 19663..19772,20332..20463))

exon complement (33880, .34038)
/gene="AT4g12430"
/number=6
intron complement (34039, .34111)
/gene="AT4g12430"
/number=6
exon complement (34112, .34219)
/gene="AT4g12430"
/number=7
intron complement (34220, .34294)
/gene="AT4g12430"
/number=7
exon complement (34295, .34366)
/gene="AT4g12430"
/number=8
intron complement (34367, .34468)
/gene="AT4g12430"
/number=8

Alignment Scores:

Pred. No.: 1,1e-12 length: 137519
Score: 573.50 Matches: 240
Percent Similarity: 35.37% Conservative: 90
Best Local Similarity: 25.72% Mismatches: 274
Query Match: 11.95% Indels: 330
Gaps: 36

US-09-920-705-3 (1-901) x AT11P17 (1-137519)

QY 145 GlnAlaIySerSerMetAengIyLysProMetGlyGlySerPheLeuArgValAspPheLeu 164
DB 132829 CAGGAAAAGTCATCAACTGATGATCAGCATGAT-----GATATATAT 132870
QY 165 ArgSerGlnAlaProLysLysGlnGlnTPrAlaGlySerTrpAspAsnArgAengIyAsn 184
DB 132871 CGCATGATGATGACAGCGGTCTGCAGCAAGAGGTCGCTTTGTTCMAAGGAGCTCAGA 132930
QY 185 MetAsnHisLysProGlnTrpProHisSerTrpGlnAspPheLys-----GlyAsp 201
DB 132931 ATGCGTTAT-----GAGACCCAGACACATATATGCAATCAAAATGAAATGATAGAAAT 132984
QY 202 ValGlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAsp 221
DB 132985 GCGAAACCCAGTACGATGATATATATGAGGTTTCT-----CCTTCGTTGAAAGTATGAT 133038
QY 222 GlnGlnIleLeuHisAsnAlaMetIleLeuPheGlyGlnIleGlnArgValLysSerTrp 241
DB 133039 GATGCGCTCTTGAGGAACGCTTTCTTCGTTTGAGAAATTAACAAGGTCAGGTAATTC 133098
QY 242 ProSerTrpAsnPheAlaLeuValGlnPheArgSerAlaGlnGlnAlaArgGlnCysLys 261
DB 133099 CCGGTCGCGCATGATGATTTGTTCAATTCGAAATCTGATGCGAGCTTGTAAGGGGAAA 133158
QY 262 GlnGlyLeuGlnGlyLysTrpLeuPheAsnAsnProArgIleLysIleMetTrpSerAsnAsp 281
DB 133159 GAAAGCTTCAGGAAAGTTATTTGGCAATCTCGATGCAATATTTGTTTGCAAAAGAGT 133218
QY 282 Gln-----LeuProGln 286
DB 133219 GAACCTTCCTCATCTGCGACATGGAAGGGTCGCGGAGTATGATCATCTCTCCACCTTAC 133278
QY 287 Gln-----AspAspHisSerPheTrpSerGlyMetLysArgSerArgThr---AspMet 303
DB 133279 AGATCTGTGATGACTCGGATCTTCAGAAAGTTATCTTCAGATAGAAACTATGAGAGC 133338
QY 304 PheAsnAsnAspProSerCysValSerSerProHis----- 315
DB 133339 ATCAGCAGATTTCCCACT---GTGAGGGAACCATTAATCACTAAGATAGGAGATTAGAA 133395
QY 316 -----SerThrGlyIleProGly 321
DB 133396 GATTCTGAAGCTTACATTTTAAACAGGAACGAGACTCAAGTATGCGATGAGGTCCTGCA 133455

QY 322 SerMetArgProLeuArgGlyThrAengIyLysSerTrpAsnGlyValaGlnTrpAsnAsp 341
DB 133456 TATGGAAGTTCG---AGATCTACGCAATGATTTCTCCAGAGATATGCAATATATCTGCT 133512
QY 342 -ValValGlyLysGlnProAsnTrp-----ArgArgProSerAla 354
DB 133513 AGTCTCGGGGAAA-----TGGGTACTTCATTCGCGATATATCTCAGAGATTCAG 133563
QY 354 aAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaI 374
DB 133564 ACAAGAG-----GTTCAAGATATGAAAGAGCCATGGGATCTACCCAGAGATGACTAC 133614
QY 374 nGlyThrArg-----ArgProMetArgSerAsnProAspSerTrpGlnGlyTrpAspPr 392
DB 133615 TACTACCAAGGAAATCAGAGATGGAAGACAGATCTCCAGCCTGAGAGGAGCTTC-CC 133673
QY 392 oAla---GlnLeu-----ValArgLysSerLysArgThrArgArgAspGlySerValAs 409
DB 133674 AGGGCATCAGCTTTTCGTATGAGAGCAAGAAAGAGCCCTTTCAAGGGCATCTGCTGA 133723
QY 409 pGlyPheThrProMetGlyValAspGlnArgSerPheGlyArgGly----- 424
DB 133734 T---TTTCCCAAAAGATGATTCATTCAGCGGAACATATGAGGCTGGACAGCTAAGATCAA 133790
QY 424 ----- 424
DB 133791 CCAACAGTTGACGACCGCTTAACCTGGCTATTAAGAAATGGGACACAGATGACTTAC 133850
QY 425 -----SerValAlaAlaArgProI1 431
DB 133851 AGAACCATGAAAGATTATAGGAGGATTATCTCTGCATCAGTTGTCCTGAAAGAAA 133910
QY 431 eArgGlyProProAsp-----SerAspHisIleTrpArgGlyMetI1 445
DB 133911 AAGGTACACCCCTGACATGATGACACATCACTAAAGAGCTGAACTGGGAAGGACTAT 133970
QY 445 eAlaLysGlyGlyThrProValCysCysAlaArgCysValProMetGlyLysGlyIleG1 465
DB 133971 TGTCTAAGGAGAGCAATCCATTTGTGACAAATATGCTTTCTGTGGGAAAGTATGGA 134030
QY 465 uThrLysLeu----- 468
DB 134031 CATGATGCTGTAAAGTACGCTTAAGCTTATTTCTGTATTTCTGTTTGAATGGCTC 134090
QY 468 ----- 468
DB 134091 TCTATATTTGCTTAAGTATTCAATCAACAAAGTATAGTCAATCCGTTTGTATTTGTT 134150
QY 468 ----- 468
DB 134151 TAAGACTTTTCTTAATTCAATGTTCTTTTGTGTATGCTTTAGACTTGTATCAAA 134210
QY 469 -----ProGlnValAlaAsnCysSerAl 476
DB 134211 TTAATTTTATGTTTGTATGTTATATAAGCAGGCTGAGTTTCTATGATTTGACGCGC 134270
QY 476 aArgThrAspLeuAsnMetLeuAlaLysHisTrpAlaValAlaIleGlyCysGlnLysLe 496
DB 134271 AAGAACTGGTTTAAAGCATGCTGCTAAGCATTAACCAATCATCTAAAGCATGGGTGCT 134330
QY 496 lPhePheValProAspArgGlnGlnAspPheAlaSerTrpThrGlnPheLeuArgTrpLe 516
DB 134331 TTTCTTTGTTCTCGAAGTGAATGCTGATATCGTGTATTATATGATGATTATGATATCT 134390
QY 516 uSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuVal 536
DB 134391 GGAAGAGAGCAACGGGAGCGCTTTCTTAATTTGATGACACAAACAGCTTTCTCTGCT 134450
QY 536 lProProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGlnArgLeuTrpL 556
DB 134451 GCTTCATCTGATTTCTCCGAGAAAGTACTTAAGTTCTCTGGGAAACATTAAGCATCTCGG 134510
QY 556 yValValLeuLysLeuProProAlaValProValThrAlaSerTrpArgGlnGlnLys 576

Job time : 10611 secs

```
Db 134511 AGTATTCTACGTTA----- 134526
QY 576 rGlnSerAsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHis 596
Db 134527 -----GAAATGGTGGTCTGG 134543
QY 596 sSerLeuTyrProProArgGlnuAntyTyrIleArgGlyAlaProGlnuHisLeuThrAlaAl 616
Db 134544 CTCGGGCGGGGCTGTTCAACAGCAAGGTGAGAGAAAAGATACAGACTTGCTTAATTATTA 134603
QY 616 aSerLysProSerValSerGlnProLeuArgIleProAsnAsnAlaAlaProGlnAlaG 636
Db 134604 TGGTGAACAATCATATTCAGAA-----GCTAGTGAAGCGTTCTCGATGTCGG 134651
QY 636 yValSerLeuThrProGlnuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerG 656
Db 134652 A-----AA 134654
QY 656 nProAlaAlaProGlnuSerHisGlnProMetSerGlyProSerThrValValSerThrAl 676
Db 134655 CCCACGATATCCG-----GGTCCACAGCTTTCTGCGTAGTGC 134693
QY 676 aHisGlnSerAsnGlyLeuTyrAsnGlyGlnAlaProSerGlnAlaIleTyrLysArgGlyPr 696
Db 134694 GGGTCGTGACAAAT-----CAAAAGTCATCAATGATCC 134726
QY 696 oGlnThrValHisAspAlaSerAsnGlnuSerPheGlnGlnTyrGlyAsnGlnTyrThrPr 716
Db 134727 ATAC-----GTGAGAGACAGCATATCATGTTGAGTCAACCGGATTTCTGA----- 134772
QY 716 oAlaGlyGlnLeuProProProProSerArgTyrProProAlaSerAsnAsnProAsnTyr 736
Db 134773 -----TCGAGCTGGCCGCTCGATACAAACCCAGAG 134807
QY 736 rThr-----SerGlyMetValHisGlyAsnMetG 746
Db 134808 TTACACCTTCATAGATCATACTGTTCAAAAACACAGCGGGTTGTTCTCGGGAACAGCA 134867
QY 746 nTyrGlnSerGlnSerValAsnMetProGlnuLeuSerProLeuProAsnMetProHisAs 766
Db 134868 A-----AATGC 134873
QY 766 nAsnTyrSerMetTyrThrGlnGlySerSerAsnHisProValSerGlnProMetValG 786
Db 134874 AGATCTAAGTCGTAT-----CATGATACGAAATCCAGTCCAGC 134915
QY 786 nGlnTyrGlnProGln-----AlaSerMetProAsnGlnAsnTyrG 800
Db 134916 AGGGTTTCAGCTGAACAGCTTACACATTAGCTTCTCGTACCCAGACAGCAA----- 134970
QY 800 yProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGlnAlaGlnAs 820
Db 134971 -----CAACAGGTACAG-----AACACTCCAAATCAGCTGAGAG 135005
QY 820 nLeuAsnPro-----SerGlnPheGlnAlaAlaMetGlnProProAl 834
Db 135006 ATATGCACCAAGAGCCGAGCAAGTTTCAGCATTTGCAACATGCG--CAGACACCAAG 135062
QY 834 aAspLysAlaAsnLeuGlnuProGlnuAsnGlnAlaLeuArgLeuGlnProMetIleSerG 854
Db 135063 CATACTCTCAATTGTTACTTCCCAAAATCAAAATGATACAGATACAA-----AGTAG 135113
QY 854 yAspGlyGlnGlyThrThrAspGlyGlyValAlaAspLysAsnGlnArgTyrGlnSerThre 874
Db 135114 CAACAGCCCAACAAGAGAGAGAGGCTAATCCACAGAAAGCGTTTCAAGCAACATT 135173
QY 874 uGlnPheAlaAlaAsnLeuLeuGlnIleGlnGln 886
Db 135174 ACAGCTAGCAGCTGCACCTTCTCCAGCAGATTCACAA 135210
```

Search completed: April 8, 2005, 12:01:29

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd

OM protein - 'nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 07:37:17 ; Search time 1152 Seconds

4629.937 Million cell updates/sec

Title: US-09-920-705-3

Sequence: 1 MALSMKPFRAADSGFQSNL.....LQIQKQKQSSGTPAGGP 901

Scoring table: BLOSUM62

Ygapop 10.0 , Ygapext 0.5

Deleop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

```
Minimum DB seq length: 0
```

[illegible]

Maximum Match 100%

2017-18

```
-MODEL=frame+2p2n.model
```

```
-DB=N_Geneseq -QFMT=fastap -SUFFIX=mg -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
```

-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL

```
-USER=US09920705 @CGN_1_1_581 @runat_07042005_101812_20618 -NCPU=6 -ICPU=3
```

```
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
```

1: geneSeq19808:4

3: geneSeqn2000s: *

5: geneBegin2001bs:

7: geneSeqn2002bs:

9: geneseqn2003bs:

11: geneSeqn2003ds

13: geneSeqn2004b

Pred. No. is the number of re

and is derived by analysis of

Result	Query
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

2	4656.5	97.0	3171	12	7
---	--------	------	------	----	---

C	4	1436	29.9	801	6	AE
---	---	------	------	-----	---	----

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4801	100.0	2706	6	ABA95167	ABA95167 Arabidops
2	4656.5	97.0	3171	12	ADN73104	ADN73104 Thale cr
3	4392	91.5	4593	6	ABA95166	ABA95166 Arabidops
4	1336	23.9	801	6	ABA95168	ABA95168 Arabidops
5	651.5	13.6	3715	6	ABA95170	ABA95170 Arabidops

5	466	9.7	349	6	ABA95169	AB095169 Arabidops
7	336.5	7.0	6836	6	ABZ21692	ABZ21692 Human RBM
8	302	6.3	2190	3	AAZ51763	AAZ51763 Human RNA
9	302	6.3	3127	13	ACN31956	ACN31956 Yimour-as
10	302	6.3	87980	8	AA053223	AA053223 4
11	297	6.2	2673	13	AD595035	AD595035 5 of
12	297	6.2	2673	13	AD595197	AD595197 Human the
13	287.5	6.0	3084	4	AB105595	AB095197 Human RBM
14	287.5	6.0	9744	4	AB105594	AB105594 Drosophi
15	276.5	5.8	3383	8	ABZ21697	ABZ21697 Human RBM
16	273	5.7	3192	4	AAH17592	AAH17592 Human CDN
17	273	5.7	3312	8	ABZ21695	ABZ21695 Human RBM
18	273	5.7	3423	8	ABZ21696	ABZ21696 Human RBM
19	271.5	5.7	10656	12	AD000904	AD000904 Mouse hom
20	268	5.6	3288	6	ABZ11870	ABZ11870 Human pol
21	268	5.6	3228	12	ADMA4388	ADMA4388 Novel hun
22	265	5.5	10995	12	AD000903	AD000903 Human hom
23	265	5.5	11145	8	ACD13385	ACD13385 Human DNA
24	265	5.5	11680	3	ACC67700	ACC67700 Human ORF
25	265	5.5	12227	8	ACC50291	ACC50291 Breast ca
26	265	5.5	12227	12	ADK60461	ADK60461 Angiogen
27	265	5.5	12227	12	ADK60762	ADK60762 Angiogen
28	265	5.5	12227	12	ADP73084	ADP73084 Angiogen
29	265	5.5	12227	13	AD588605	AD588605 Human hou
30	265	5.5	12227	13	ADP23327	ADP23327 PRO polyIV
31	261	5.4	2505	13	ADRO7405	ADRO7405 Full leng
32	261	5.4	11295	5	AD145263	AD145263 Human ova
33	254	5.3	438	6	AB194207	AB194207 Arabidops
34	247	5.1	3046	4	AB109397	AB109397 Drosophi
35	247	5.1	3389	4	AB109385	AB109385 Drosophi
36	246.5	5.1	1436	6	ADP81563	AD181563 Leukemia
37	246.5	5.1	1477	12	ADL13065	ADL13065 Human stem
38	246.5	5.1	1871	4	AA661035	AA661035 Human can
39	246	5.1	2228	13	ADR24549	ADR24549 Breast ca
40	246	5.1	2482	12	ADQ23974	ADQ23974 Human so
41	243	5.1	2228	2	AAQ34554	AAQ34554 CDNA enco
42	240	5.0	1766	6	AB199779	AB199779 Mouse isc
43	236.5	4.9	8147	6	AB873113	AB673113 DNA encod
44	236.5	4.9	8147	10	ADK11302	ADK11302 Human CRR
45	236.5	4.9	8677	12	ADQ85476	ADQ85476 Human tun

ALIGNMENTS

Accession	Gene	Species	Location	Qualifiers
AB095167	standard	CDNA	2706 BP	
AB095167				
AB095167				
20-MAY-2002	(first entry)			
Arabidopsis floral induction gene (FPA) cDNA sequence.				
Arabidopsis floral induction; photoperiod; plant; flowering; FLC; FRI; frigid;				
flowering locus C; transgenic; gene; ss.				
Arabidopsis thaliana.				
Location/Qualifiers				
1..2706				
/*tag= a				
/product= "FPA protein"				
46..279				
/*tag= b				
/note= "RNA binding region"				
283..522				
/*tag= c				
/note= "RNA binding region"				
610..852				
/*tag= d				
/note= "RNA binding region"				

PN WO200212518-A2.
 XX 14-FEB-2002.
 PD
 PF 02-AUG-2001; 2001WO-US024427.
 XX 03-AUG-2000; 2000US-0222550P.
 PR
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 PA
 XX Amasino RM, Schomburg FM, Michaels SD, Patton D;
 PI WPI: 2002-227160/28.
 DR P-PSDB; ABB07659.
 XX
 PT Novel isolated DNA sequence comprising coding sequence for floral
 induction gene (FPA) gene which controls flowering time in plants, useful
 for altering flowering time of plants and for downregulating flowering
 locus C (FLC) mRNA activity.
 PT
 XX
 PS Claim 5; Page 29-34; 39pp; English.
 CC The invention relates to the FPA gene (floral induction promoter in
 CC plants during both long and short day photoperiods) from Arabidopsis
 CC thaliana. The FPA coding sequence is useful for altering flowering time
 CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
 CC activity. The FPA gene can accelerate flowering in several late-flowering
 CC mutant backgrounds and can fully compensate for addition of two naturally
 CC occurring genes, *frigida* (FRI) and FLC which confer late-flowering
 CC phenotypes. Over expression of FPA can compensate for the delaying effect
 CC caused by short days on floral induction, and decreases FLC mRNA in
 CC plants containing the FLC gene. Fragments of the FPA gene can also act to
 CC decrease activity of an endogenous FPA gene by modifying the expression
 CC of the endogenous FPA gene and expression of a portion of polypeptide
 CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
 CC present sequence represents the A. thaliana FPA protein encoding cDNA
 XX
 SQ Sequence 2706 BP; 779 A; 628 C; 617 G; 682 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3,74e-246 Length: 2706
 Score: 4801.00 Matches: 901
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-09-920-705-3 (1-901) x ABA95167 (1-2706)
 QY 1 MetalAlaLeuSerMetIysProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
 DB 1 ATGGCGTTATCTATGAGCCATTCAGAGCGATGATTCGCGTTTCAGTCAAAACATCTT 60
 QY 21 TTPValGlySerLeuThrProGluThrThiGlnSerAspLeuThrGluLeuPheGlyArg 40
 DB 61 TGGGTCGTAGACCTTAACGCCGAGACGACAGAGTCAGATCTGACGAGTTGTTGAAGA 120
 QY 41 TTYGlyAspGlnLeuAspArgGlyLeuThrValTyrSerSerArgGlyPheAlaPheIleTyr 60
 DB 121 TACGGGATATTGATGAGATCAACGGTGTATCTTACACAGAGGCTTTCCTTATATATCAAC 180
 QY 61 ArgHISvalGluGluAlaValAlaIalysGlnIalysGlnGlyAlaAsnLeuAsnGly 80
 DB 181 ACACATGTGGAG 240
 QY 81 SerGlnIleLeuValIleGluTyrAlaArgProAlaIysProCysIysSerLeuTyrValGly 100
 DB 241 AGTCAAAATTAAATGAGATACGACGACGCGGAAAACCTTGTAAAGATCTATGGGTGGGT 300
 QY 101 GilyIleGlyProAsnValSerIysAspAspLeuGluGluGluPheSerLeuPheGlyIys 120
 DB 301 GGAATCGGCGCTTAATGTCTCCAAAGGATACCTGGAGAGAGATTCAAGCAAGTTTGGGAAA 360

QY 121 IIEGluAspPheArgPheLeuArgGluArgIysThrAlaPheIleAspTyrTyrGluMet 140
 DB 361 ATCGAGGATTTTAAAGTTTCTCAAGAGACCAAGACGCTTCATTGATTTATATAGATG 420
 QY 141 AspAspAlaLeuGlnAlaIysSerMetAsnGlyIysProMetGlyIysSerPheLeuArg 160
 DB 421 GATGATGCTTTACAGGCTTAAGAGCATGATGAGAAAGCCTAATGGGTGTAGCTTTTGGCT 480
 QY 161 ValAspPheLeuArgSerGlnAlaProIysIysGluGlnTTPAlaGlySerTyrAspAsn 180
 DB 481 GTTGAATTTCTCCGCTCACAGCGCCAAAAGAAAGACATGGCGCTCTTACGATAC 540
 QY 181 ArgAsnGlyAspMetAsnHisIysProGlnTyrProHisSerTyrGluAspPheIysGly 200
 DB 541 AGAAATGGCAATATATATCATTAACCGCAGATATCTCACTCATATGAAAGCTTTAAAGA 600
 QY 201 AspValGlnProSerIysValLeuThrIleGlyPheProProThrAlaThrGlnCysAsn 220
 DB 601 GATTCACAGCCAGTAAAGGTTCTGTGATGGGTTCCTCTACTGCTACCAATGCAAT 660
 QY 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValIysSer 240
 DB 661 GATGAGCAAAATTTCTGCACAATGCCATGATATCTTTGGTGAATCGAGAGGTAAAGT 720
 QY 241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
 DB 721 TACCATCAAGAAATTTTGACCTGTGAGATTGAGAGCGCGGAGAGAGCTCCGCAATGC 780
 QY 261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleIysIleMetIysSerAsn 280
 DB 781 AAGGAAGGCTTACAGGGAGGAGTTTTCATTAATCTTGAAATCAAAATTAATGATCTCAAC 840
 QY 281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetIysArgSerArg 300
 DB 841 GATAGTTGCTCTCTGAGCAAGACATGATTAATCTGATTAATGATGAACGGTCAAG 900
 QY 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
 DB 901 ACAGATATGTTCAATATGATCTTCAATGATGATCTTCTCAATCTTCTCAATCTTCAATCTT 960
 QY 321 GlySerMetArgProLeuArgGlyTyrAsnGluIysSerTyrAsnGlyAlaGluTyrAsn 340
 DB 961 GGGTCTATGAGGCCCTCAAGAGGTACGAATGAGCGTTCAATATAGTGCAGATACAT 1020
 QY 341 AspValAlaGlyIysGluProAsnThrPheArgProSerAlaAsnGlyThrGlyIleLeu 360
 DB 1021 GACGTTGTGTAAAGAGCCAACTGAGAGAGCCATCTGCAAAATGAACTGGAAATATCTC 1080
 QY 361 ProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMet 380
 DB 1081 CCATCTCCAAACAGGAGCTGGAATCTTCCATCTCTGCAACAGGTACGAGCGCCCTTAATG 1140
 QY 381 ArgSerAsnProAspSerTyrGluGlyTyrAspProAlaGlnLeuValArgIysSerIys 400
 DB 1141 AGGTCAAACCCCATTTCTGGGAAGATATGATCTCTGCTGATGGTGTACAGAAAGTAAA 1200
 QY 401 ArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGluArgSer 420
 DB 1201 CGAACAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
 QY 421 PheGlyArgGlySerValAlaIalysProIleArgGlyProProAspSerAspHisIle 440
 DB 1261 TTTGGTCGAGGTTCAATGCTGCTGATACCTATCCGAGGCGCCCTGATTTCTGATCACTA 1320
 QY 441 TTPArgGlyMetIleAlaIysGlyIysThrProValCysCysAlaArgCysValProMet 460
 DB 1321 TGGAGAGATGATTTCCAAAGGTGAACTCCGCTGTGTGCTGCTGTGTGCTGCTGCTGCT 1380
 QY 461 GilyIysGlyIleGluThrIysLeuProGluValAlaAsnCysSerAlaArgThrAspLeu 480
 DB 1381 GGAAGAGGATGGAATCAATCACTGAGTGCATGATTTGTCAGCAAGAACTATTTTG 1440
 QY 481 AsnMetLeuAlaIysHisTyrAlaValAlaIleGlyCysGluIleValPhePheValPro 500

Db	1441	AATATGCTCGGTAACAATTACGCCGCTCCANTGGATGTGATGATGTTTTTTTGGTACCA	1500
Oy	501	AspArgGluGluAspPheAlaSerThrThrGluPheLeuArgThrLeuSerSerIysAsp	520
Db	1501	GACAGGGAGAAAGATTTTGGCTGCTTACCTGAATTTCTCCGGTACCTTAGCTCAAAAGAT	1560
Oy	521	ArgAlaGlyValAlaIysLeuAspAspGlyThrThrLeuPheLeuValProProSerSerAsp	540
Db	1561	CGGGGGGGGTGGTCCCAAAATTGATGATGATGATCAAACTTTATTTCTGGTGCCCTCCATCAAT	1620
Oy	541	PheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuThrGlyValIValLeuIys	560
Db	1621	TTCTTTAACTGATGATCTCCAACTGACCCGCTCAAGAACGGCTATATGTGTGTTCTCAAG	1680
Oy	561	LeuProProProAlaIValProValThrIlaSerThrArgGlnIysSerGlnSerAsnPro	580
Db	1681	TTACCCCGGCGAGCGGCTCTGTATCAGCATCATCAGACAAGATCTCACTGCTCAATCTCT	1740
Oy	581	LeuHisThrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuThrPro	600
Db	1741	CTGCATTATATGATGATCAAGCCCGGGATTCACCTGCAATGCTATGATCAAGTTTATATCT	1800
Oy	601	ProArgGluAsnThrIlaArgGlyValArgProGlnHisLeuThrAlaIaSerIysProSer	620
Db	1801	CTTAGGGAAATTTACATTATGGGGGTGCACAGACATTTGACAGCTGCTTCAAAACATCT	1860
Oy	621	ValSerGluProLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr	640
Db	1861	GTTAGCGAGCCCTCAGAAATACCTAATATGAGAGCGCTCAAGCGGGGTTATGTTTAACT	1920
Oy	641	ProGluLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaIaIaPro	660
Db	1921	CCGGAGCTTTTAAAGCACCTCTGGCATCTATTTCTCCCTGCACTTTCACACTGCTGCCCT	1980
Oy	661	GluSerHisGlnProMetSerGlyProSerThrValIaSerThrAlaHisGlnSerAsn	680
Db	1981	GAGATCACCAACCTATGATCAGAGACTTTCACAGATGTTTTCACAGCACATCAGTCCAT	2040
Oy	681	GlyLeuThrArgGlnGlyAlaIaProSerGlnAlaIaTrpIysArgGlyProGlnThrValHis	700
Db	2041	GGAGCTGTCATGAGAGAACACCGCTCAAGCTTGGAAAGAGGTCCCAACCAACTGAT	2100
Oy	701	AspAlaSerAsnGlnSerPheGlnGlnThrGlyAsnGlnThrProAlaGlyLeuLeu	720
Db	2101	GATGGGTCAAAATCAGTCAATCCAAATATCGGAATTCAGTACATCCAGCTGGGCACTA	2160
Oy	721	ProProProProSerArgThrProProAlaSerAsnAsnProAsnThrThrSerGlyMet	740
Db	2161	CCTCCTCCTCTTCCGCTTACCTCCACCTTCAACCAACCCCACTACATCAGTGGAGATG	2220
Oy	741	ValHisGlyAsnMetGlnThrGlnSerGlnSerValAsnMetProGlnLeuSerProLeu	760
Db	2221	GTCATGGCAATCAGTCAATACCAAGGCCAATCTGTTAACTGCTCAGCTGTCTCCGTTA	2280
Oy	761	ProAsnMetProHisAsnAsnThrIserMetThrThrGlnIysSerSerAsnHisProVal	780
Db	2281	CCAAATATGCTCATATATATATTCATGTACACTCAGGGTGTGTCAAATCATCTGTT	2340
Oy	781	SerGlnProMetValGlnGlnThrGlnProGluAlaSerMetProAsnGlnAsnThrGly	800
Db	2341	TCTCAGCCCATAGCTCCAGCAATACCAACAGAGGTCTCAGGCCAAACAAATCATATGT	2400
Oy	801	ProIleProSerThrGlnGlnAlaAsnHisGlyValIleThrThrAsnGlnAlaGlnAsn	820
Db	2401	CCAAATTCAGATTATCAGCAAGCTATTTTCATGGCGTAAACAAATCAGGACACAGAC	2460
Oy	821	LeuAsnProSerGlnPheGlnAlaIaIaMetGlnProProAlaAspIysAlaAsnLeuGlu	840
Db	2461	TTAAACCCCTTCCCAATTTCAAGCTGCATGCAACCAACGACGAGATTAAGCAAAATTTAG	2520
Oy	841	ProGlnAsnGlnAlaIaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThr	860

Dd	2521	CCACAAAACCAAGCACAATGCGTTCAGCTGTATCTCTGGGAGTAGTGACGGGTAAACA	2588
Oy	861	Aapglgtlvalalapylaenqlnarytyrglnserthrleuglnlphealaiaalenue	880
Dd	2581	GATGGGAGGGTCGATTGAAGAATCACAGATACCAAGTCAACACTACAAATTGCAGCAAACTT	2640
Oy	881	Leuleuglnilegnglnlntysglnlglnlgnlserserglythrproalaglvglngly	900
Dd	2641	CTTCTCCGATACAGACGAAACACAGACGACACGTTCTCAGTACTTCGGGCTCGACAGGGG	2700
Oy	901	Pro 901	
.Dd	2701	CCT 2703	
RESULT 2			
ID	ADN73104	standard; cDNA; 3171 BP.	
AC	ADN73104;		
DT	15-JUL-2004	(first entry)	
DE	Thale cress cDNA upregulated in E2Fa/Dpa expressing plants Segid 999.		
KM	gene; ss; plant; transgenic; E2Fa/Dpa transcription factor;		
KW	growth regulator; animal feed product; thale cress;		
KX	cell wall biosynthesis; nitrogen metabolism; carbon metabolism.		
OS	Arabidopsis thaliana.		
XN	WO2004035798-A2.		
PN	29-APR-2004.		
PD	20-OCT-2003; 2003WO-EP011658.		
PF	18-OCT-2002; 2002EP-00079408.		
PR	(CROP-) CROPPDESIGN NV.		
PA	Inze D, De Veylder L, Vlieghe K;		
Pt	WI: 2004-348466/32.		
DR	P-PSDB; ADN73105.		
PT	Altering plant characteristics, useful for producing plants for enzyme or pharmaceutical production comprises modifying in a plant, expression of one or more nucleic acids and/or modifying level or activity of one or more proteins.		
XX	Claim 1; SEQ ID NO 999; 134pp; English.		
XX	This invention relates to a novel method for altering one or more plant characteristics. Specifically, it refers to identifying genes that are up - or down-regulated in transgenic plants overexpressing the heterodimeric E2Fa/Dpa transcription factor of Arabidopsis and using these sequences to alter plant characteristics accordingly. The present invention describes generating transgenic plants for the production of growth regulators, enzymes, therapeutics, pharmaceuticals and animal feed products, where the altered plant characteristics are selected from increased yield or biomass, enhanced survival capacity, stress tolerance, plant architecture or physiology, altered lipid mobilisation and/or altered photosynthesis, transduction, storage lipid mobilisation and/or altered photoassimilation, each relative to the corresponding wild type plants. Accordingly, these sequences can also be useful as positive or negative selectable markers during transformation of cells or tissues. The identified genes play a role in a variety of biological processes such as DNA replication, cell wall biosynthesis, nitrogen and/or carbon metabolism or they function as transcription factors. This polynucleotide sequence is thale cress cDNA upregulated 1.3 fold or more in plants overexpressing the E2Fa/Dpa transcription factor, given in an exemplification of the invention.		
Sequence	3171BP; 910 A; 723 C; 709 G; 829 T; 0 U; 0 Other;		

Alignment Scores:

Pred. No.:	2,13e-238	Length:	3171
Score:	4656.50	Matches:	892
Percent Similarity:	85.04%	Conservative:	6
Best Local Similarity:	84.47%	Mismatches:	3
Query Match:	96.99%	Indels:	155
DB:	12	Gaps:	3

US-09-920-705-3 (1-901) x MDN73104 (1-3171)

```
QY      1 Mech1aleuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
Db      1 ATGGCGTTATCTATGAAAGCCATTCAGAGCCAGATGATCCGGTTTCAGTCAAAACATCTT 60
QY      21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
Db      61 TGGGTCGGTAGCCTTAACGCCGAGACGACAGATCTGACCGAGTTGGTGGAGA 120
QY      41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
Db      121 TAGCGCATATGTATGATCAAGGTATTTCTTCAAGAGCTTTCGGTTATATATCTAC 180
QY      61 ArgHisValGluGluValAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
Db      181 AGACATGTGAGAGAGAGACAGTCGACGCCAAAGAGGCTCTTCAGAGCAAAATTTGAATGA 240
QY      81 Ser-----
Db      241 ACGAATGAGCTCTGCAAAAGAAATACATCAAAACATGTTTACTCTCGTTTATGCCAGAA 300
QY      81 -----
Db      301 GGTTTAATCCGATTGCTTATATCTGCAAACTGTATCCGCGTTTAGCAGTCTGATTCA 360
QY      81 -----
Db      361 GGTTCCTTCAGATGTTAAATCTCACAGCTTGAGTGAATGATATGCGTACTCCACTC 420
QY      82 -----
Db      421 GAAATTCGCAATGATAAAGAAATTTTGGAGTTTAAAGTTCAAGTTTCTTCCGAAAAATCAAA 480
QY      87 -----
Db      481 CCGCGTGAAGAAATTTGTCTATGCGCGTACGTAGATATCATATACATGATCGCAACA 540
QY      87 -----
Db      541 GAAATTTCTGGCAACCGTATCGTCTAATCTCTAGTATCGTGGCACATATCCATATCAG 600
QY      87 -----
Db      601 TGTGAGAGAGTTTGGGCTTGTGTTTGGCTGCATGCTTCGAGAAAGCAAAATTACGTGGTTG 660
QY      87 -----
Db      661 TTAAGCGGCTAATCTTACCAATACGAAACGCTTATTCGAGAAACCATGTTTGTCTCC 720
QY      88 ---AlaArgPro---AlaLysProCysLysSerLeuThrValGlyTyrIleGlyProAsn 105
Db      721 AATTCACATCCCTTACGCAAAACCTTGTAAAGATCTATGGGATGGGAAATCGCCCTAT 780
QY      106 ValSerLysAspAspLeuGluGluLupheserLysPheGlyLysIleGluAspPheArg 125
Db      781 GTCTCCAAAGATGACCTGAGAGAGAGTTCAACAAAGTTTGGGAAATAGAGATTTTACG 840
QY      126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
Db      841 TTTCCTAGAGAAAGCAAGACAGCTTTCATATGATTTATGAGATGAGATGAGATGCTTACAG 900
QY      146 AlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeuArg 165
```

```
Db      901 GCTAAGACATGATGAAAGCCCTATGGGTGTAACCTTTTGGCTTGATTTTCTCCGG 960
QY      166 SerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArgAsnGlyAsnMet 185
Db      961 TCACAAAGCCCAAAAGAAAGCAATGGCTGGCTCTTACGATTAACAGAAATGGCAATATG 1020
QY      186 AsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspValGlnProSer 205
Db      1021 AATCATTAACCGCAGATATCTCATCATATGAAAGCTTTAAAGAGATGTCAGCAAGT 1080
QY      206 LysValIleuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeu 225
Db      1081 AAGTTCTGTGATGGATGGGTTCCCTCTCTCTCTCATCATCATGATGATGACAAATTTTG 1140
QY      226 HisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsn 245
Db      1141 CACAATGGCATGATATCTCTTGGTGAAGATCGAGAGGGTAAAGATTACCATCAAGAAAT 1200
QY      246 PheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluGlyLeuGln 265
Db      1201 TTTCGACTTGTGAGATTGAGAGCGGAGAGAGCTGCCAATGCAGAAAGGCTTACG 1260
QY      266 GlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProPro 285
Db      1261 GGAAGTTATTCAATTAATCCTAAGATCAAAATTAATGATCTAAACGATGAGTGCCTCT 1320
QY      286 GluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsn 305
Db      1321 GAGCAAGACGATACGATGTTTACTCTGATGAAAGCGTCAGCAAGACGATATCTTCAAT 1380
QY      306 AsnAspProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgPro 325
Db      1381 AATGATCTTCAATTTGATATCTTCTCTCATTTCTATGAAATTCCTGGGCTTATGAGGCC 1440
QY      326 LeuArgGlyThrAsnGluArgSerTyrAsnGlyValaGluTyrAsnAspValaGlyLys 345
Db      1441 CTCAGAGGTACAAATGAGCGTTCAATATATGTCAGAAATACATATGACGTTGTTGGTAAG 1500
QY      346 GluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGly 365
Db      1501 GAGCCAAACGTGAGAGAGCCATCTGCAAAATGGAACGTGAATATCTCCATCTCCAAACGGA 1560
QY      366 ProGlyIleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAsp 385
Db      1561 CTTGGAATCCCTCCATCTCTCTGCAAGATACAGAGGCGCTTATGAGGTCAAAACCCGAT 1620
QY      386 SerTrpGluGlyTyrAspProAlaGlnLeuValArgGlnSerLysArgThrArgArgAsp 405
Db      1621 TCTTGGAGAGATATGATCTCTGCTCACTGCTGTCAGAAAGTAAACGAACCAAGAGAT 1680
QY      406 GlySerValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySer 425
Db      1681 GGATCAAGTGAAGGTTTACTCCAAATGGGTGTCATGAGAGGTCATTTGGTCAAGGTCA 1740
QY      426 ValAlaAlaArgProIleArgGlyProProAspSerAspHisIleTrpArgGlyMetIle 445
Db      1741 GTTCTGCTAGACCTATCCGATGCGGCCCTCGATTTGATACATATGAGAGAAATGATT 1800
QY      446 AlaLysGlyGlyThrProValCysArgAlaArgCysValProMetGlyLysGlyIleGlu 465
Db      1801 GCCAAGGAGTAACTCCCGTCTGTGTGCTCGTGTGTGATCTATGCGAAAGGGGATTTGAA 1860
QY      466 ThrLysLeuProGluValValaAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLys 485
Db      1861 ACTTAATGCTGCTAAGAGTCGTCATTTGTTTACGCAAGAACGATTTGAATATGCTCGCTAAA 1920
QY      486 HisTyrAlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluGluAsp 505
Db      1921 CATTAAGCCGCTGCTCAATGATGAGATGAGATGTTTTTTTGTGTCACAAACAGGGAAGAAAT 1980
QY      506 PheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAla 525
Db      1981 TTTCGCTTAACTGAATTTCTCCGGTACCTTAAGCTCAAAAGATCGGCGGTGTTGGCC 2040
```

QY 526 LyeLeuaspapglYhThrLeuPheLeuValProProSerAapPheLeuThraaVal 545
 Db 2041 AAATTAAGATGATGTAACCTTAATCTTGCTGCTCAAGATTTCTTAACCTGATGA 2100
 QY 546 LeuGlnValThraGlnGlnIuaArgLeuTgYValValLeuYsLeuProProProAla 565
 Db 2101 CTCGAAGTGAACCGCTCAAGAACGGCTATAGTGTGTTGTTCAAGTTAACCCCGCCAGCC 2160
 QY 566 ValProValThraLaseTyrArgGlnGlnSerGlnSerAandProLeuHisTyrMetAsp 585
 Db 2161 GTTCTGTATCAGCATCAACAGCAAGAACTCAGCCAACTCTGCAATATATGAT 2220
 QY 586 GlnAlaArgSerProAlaAsnAlaSerHisSerLeuTyrProProAlaGlnAantYr 605
 Db 2221 CAAGCCCGGAGATTCACCTGCAATGCTAGTCAAGATTATATCTCTTAGGAAATATAC 2280
 QY 606 IleArgGlyValaProGlnHisLeuThraAlaSerLysProSerValSerGlnProLeu 625
 Db 2281 ATTAGGGGTGCACCAAGACATTGACAGCTGCTTCAAAACCATCTGTAGCGAGCTTCTC 2340
 QY 626 ArgIleProAsnAlaAlaProGlnAlaGlyValSerLeuThrProGlnLeuLeuAla 645
 Db 2341 AGAATACCTAATATGACAGCGCTCAAGCTGGGGTTAGTTAACTCCGAGACTTTAGCC 2400
 QY 646 ThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGlnSerHisGlnPro 665
 Db 2401 ACTGTGCATCTATCTCCCTGCAACTCTCAACTGCTGCCCTGAGACTCCCAACT 2460
 QY 666 MetSerGlyProSerThraValSerThraHisGlnSerAanglyLeuYrAsnGly 685
 Db 2461 ATGTCAAGACCTTCAACAGTGTTCACAGCAGCATGCTCAATGAGACTGTCATGGA 2520
 QY 686 GlnAlaProSerGlnAlaTgYValArgGlyProGlnThraHisAspAlaSerAangln 705
 Db 2521 GAAGCACCCTCTCAAGCTTGAAAAGAGTCCCAAAACATTCATATGATGCTCAATACG 2580
 QY 706 SerPheGlnGlnTyrGlnAanglnTyrThrProAlaGlyGlnLeuProProProSer 725
 Db 2581 TCATTCACAAATACGAAATACGATCAGCTCAGCTGGGAGACTCTCTCTCTCTCG 2640
 QY 726 ArgTyrProProAlaSerAanAanProAantYrThrSerGlyMetValHisGlyAanMet 745
 Db 2641 CGTTACCTCCACCTTCAACAAACCCCACTACATGATGATGCTCATGGCAACATG 2700
 QY 746 GlnTyrGlnSerGlnSerValAanMetProGlnLeuSerProLeuProAanMetProHis 765
 Db 2701 CAATACCAAGCCCAATCTGTTAATGCTCAGCTGCTCGTTACCAATATGCTCAT 2760
 QY 766 AsnAantYrSerMetTyrThrGlnGlnYsSerSerAanHisProValSerGlnProMetVal 785
 Db 2761 AAATAATTAATTCATGTAACCTCAGGGTTCGTCAAATCATCTTTCTCAGCCCATGCTC 2820
 QY 786 GlnGlnTyrGlnProGlnAlaSerMetProAanGlnAantYrGlyProLysProSerTyr 805
 Db 2821 CAGCAATACCAACCAAGGCTCAGTCCAAACCAAACTATGCTCAATTCAGATGAT 2880
 QY 806 GlnGlnAlaAanPheHisGlyValThrThraGlnAlaGlnAanLeuAanProSerGln 825
 Db 2881 CAGCAAGCTAATTTTTCATGCGTAAACAAATCAGGCAAGAACTTAAACCTTCCCA 2940
 QY 826 PheGlnAlaAlaMetGlnProProAlaAspYValAanLeuGlnProGlnAanGlnAla 845
 Db 2941 TTTCAAGCTGCATGCAACCAAGATGAAGCAATTTTAAAGCCCAAAACCAAGCA 3000
 QY 846 LeuArgLeuGlnProMetLysSerGlyAspGlyGlnGlyThrThraAspGlyGlnValaAsp 865
 Db 3001 CTACGATTCAGACCTATGATCTCTGGGAGTGTCAAGGTCACACATGGGAGGCTCAT 3060
 QY 866 LysAanGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAanLeuLeuGlnGlnGln 885
 Db 3061 AAGAAATCAGATACCAAGCTCAACATTAATTTGACAGCAAACTTCTTCCAGATACAG 3120

QY 886 GlnYsGlnGlnGlnGlnSerGlyThrProAlaGlyGlnGlyPro 901
 Db 3121 CAGAAACAGACAGACAGCTTTCAGTACTCCGGCTGACAGGGGCT 3168
 RESULT 3
 ID ABA95166 standard; DNA; 4593 BP.
 AC ABA95166;
 DT 20-MAY-2002 (first entry)
 DE Arabidopsis floral induction gene (FPA) sequence.
 KW FPA; floral induction; photoperiod; plant; flowering; FLC; FRI; frigida;
 KW flowering locus C; transgenic; gene; ds.
 OS Arabidopsis thaliana.
 PN W0200212518-A2.
 PD 14-FEB-2002.
 PF 02-AUG-2001; 2001WO-US024427.
 PR 03-ANG-2000; 2000US-022250P.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 PI Amsino RM, Schomburg FM, Michaels SD, Patton D;
 DR WPI; 2002-227160/28.
 DR P-PSDB; ABB07659.
 PT Novel isolated DNA sequence comprising coding sequence for floral
 induction gene (FPA) gene which controls flowering time in plants, useful
 for altering flowering time of plants and for downregulating flowering
 locus C (FLC) mRNA activity.
 PS
 XX Claim 5; Page 27-29; 39pp; English.
 CC The invention relates to the FPA gene (floral induction promoter in
 CC plants during both long and short day photoperiods) from Arabidopsis
 CC thaliana. The FPA coding sequence is useful for altering flowering time
 CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
 CC activity. The FPA gene can accelerate flowering in several late-flowering
 CC mutant backgrounds and can fully compensate for addition of two naturally
 CC occurring genes, frigida (FRI) and FLC which confer late-flowering
 CC phenotypes. Over expression of FPA can compensate for the delaying effect
 CC caused by short days on floral induction, and decreases FLC mRNA in
 CC plants containing the FPA gene. Fragments of the FPA gene can also act to
 CC decrease activity of an endogenous FPA gene by modifying the expression
 CC of the endogenous FPA gene and expression of a portion of polypeptide
 CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
 CC present sequence represents the A. thaliana FPA gene sequence
 XX
 SQ Sequence 4593 BP; 1275 A; 958 C; 978 G; 1382 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3, 64e-224 Length: 4593
 Score: 4392.00 Matches: 899
 Percent Similarity: 58.72% Conservative: 0
 Best Local Similarity: 58.72% Mismatches: 2
 Query Match: 91.48% Indels: 631
 DB: 6 Gaps: 4
 US-09-920-705-3 (1-901) x ABA95166 (1-4593)
 QY 1 MetAlaLeuSerMetLysProPheAlaAspAspSerGlyPheGlnSerAanLeu 20
 Db 1 ATGGCGTTATCTATGAGCATTCAGAGCCGATGATTCGCGTTTCAGTCAAAACATCTT 60
 QY 21 TrpValGlySerLeuThrProGlnThrGlnSerAspLeuThnGlnLeuPheGlyArg 40

Db 61 TGGGTCCGTAGCTTAACGCCGAGACGACAGATCTGACCCGAGTTGTTGAGAA 120
Qy 41 TyrGluAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
Db 121 TACGGCGATATGTATTAATAATCACGGTGTATCTTCAAGAGCTTTGGCTTAAATATCTAC 180
Qy 61 ArgHisValGluGluAlaValAlaAlaLysGluValAlaLeuGlnGlyAlaLeuLysGly 80
Db 181 AGACATGTGAGAGAGACAGTCCGACCCAAAGAGCTCTTCAAGAGCAAAATTTGAATGGA 240
Qy 81 SerGlnIleLysIleGluTyrAlaArgPro----- 90
Db 241 AGTCAATTTAAGATCGAATACGACACGACCGTTTGTCTTATCTATATCTTCGTTTGTTC 300
Qy 90 ----- 90
Db 301 TCTAATCTTGATGTCTTTTGTCAAGATTAATCTCTTTTGGAAATTCATAGTCAGGT 360
Qy 90 ----- 90
Db 361 TCACAAAACCTTGATGATGCTGTGTTTAGTCAAAATAATTCCTGTGATCTGTTTTTTT 420
Qy 90 ----- 90
Db 421 TTCTCATTCGTACAAATCAAAAGTCGAAACCTAGTTTTTCTATATAGTCGTAGCTT 480
Qy 90 ----- 90
Db 481 AAGCGAAACCTGATCCGATCGAAACGTCTTTTCTCAAAATTAATCTTGTTAATTCGAACT 540
Qy 90 ----- 90
Db 541 CGCGAAAGCCAAACACAGAGAAGCTCTGCAAAATTTGATGTTAAAGCATATATACTC 600
Qy 90 ----- 90
Db 601 TTAGCAATGAGCTCTGCAAAAGAAATATACAAACATGTTTACTCTGTTTATGCGAA 660
Qy 90 ----- 90
Db 661 GAAGGTTTAATCCGATTGCTGTTTATCTGCAAACTTGATCCGCCGTTTACAGCTGATTT 720
Qy 90 ----- 90
Db 721 TCAGTTTCGCTCAGATGTTAAATCTCACAAAGCTTGATGATGATGATTTGGTACTCCA 780
Qy 90 ----- 90
Db 781 CTCGAAATTCGCAATGATGTAATAATTTGGAAGCGAGCAAACTGTATCATGCGCAAAACGA 840
Qy 90 ----- 90
Db 841 TAAGAAATTTTGAGTTTAAAGTTTCAATTCAGTTCTCTGCAAAATCAAAACCGCTGAGAAAT 900
Qy 90 ----- 90
Db 901 TTGTCTATGCGGTACGTAGATATCAATACTGATCTCGAAACGAAATTTCTGGCAA 960
Qy 90 ----- 90
Db 961 CCGTATCGTCTACTCTTAATGATGCTGGCAATATTCATATCAGTGTGAGAGATTTT 1020
Qy 90 ----- 90
Db 1021 GGGCTTGTGCTCGATGCTTCAAGAAAGCAAAATTAAGTGTGTTAAGCGGCTAAC 1080
Qy 90 ----- 90
Db 1081 TCTACCAATCAGAAACGCTTTATTCGAAAGAACATGTTTGTCTCAATTCCTCATCCCTA 1140
Qy 90 ----- 90

Db 1141 CGTACAATCTGGGCTTTTCCCATTTGATGTTCTTAAAGAAAGTGCATCTTTCACAAACATTC 1200
Qy 90 ----- 90
Db 1201 TGGGATGTGTACAGTTGACAGTGCAGACAGACAAATATTCATTTGCAGGACACTCAATTT 1260
Qy 90 ----- 90
Db 1261 CAATGCTTCCGCGCTTGAGTACTTAATGAGTGTTCAGGAGATGGTTATTTGGGAC 1320
Qy 90 ----- 90
Db 1321 TCAGCAATGATATAAAGACAGGAATGTTTGAGAAAGTATTAATGCTTCCGCTTTG 1380
Qy 90 ----- 90
Db 1381 GGGACTACATCTTCTCTACAGCAAAATCTGATGTTTAGCAAGAACTATTAACA 1440
Qy 90 ----- 90
Db 1441 CTGATTTAATAGAGAAAGAGATGTTTCTCTTCCGCAATTTATTTGTTTAGGAA 1500
Qy 90 ----- 90
Db 1501 AGCAATGATATGAATGTGTGTCGTAGTGTGAATTAATTAATTAATGTTGTGTGTTT 1560
Qy 90 ----- 90
Db 1561 AGTTCTTGATGTTGATGCTTATTTCTTGGCAACCTATCTGGGTAGTATCGCCTTCTT 1620
Qy 91 ----- 1620
Db 1621 ATCGACTTTTCTTGTGTCAGGCAAAACCTGTGAAGCTCTATGGTGGGAGAAATCGGC 1680
Qy 104 ProAsnValSerLysAspAspLeuGluGlnPheSerLysPheGlyLysIleGluAsp 123
Db 1681 CCTATGTCTCCAAAGATGACCTCGAGGAAAGTTGACCAAGTTGGGAAATCGAGGAT 1740
Qy 124 PheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
Db 1741 TTAGGTTTCTCAGAGAACGCAAGACGCTTCAATTAATTAATTAATGAGATGAGATGCT 1800
Qy 144 LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPhe 163
Db 1801 TTACAGGCTAAGAGATTAATGAAAGCTATGGGTGATGCTTTTCCGCTGTTGATTTT 1860
Qy 164 LeuArgSerGlnAlaProLysLys----- 171
Db 1861 CTCGGTCAACAAGCGCCAAAAAAGTAAGACCTCTGTGGCATTTGATTTTACTTTGA 1920
Qy 171 ----- 171
Db 1921 AAAGCTCCAGTAAACATTTTGTGTTGATTCATTAATTTGCGTCAAACTGATAGGGCTGAG 1980
Qy 171 ----- 171
Db 1981 CTCTGTCTTGCGCCCTAGAGACAGATATTACTGTCTCATTTTCATTTAGAGTAGGCT 2040
Qy 172 ----- 172
Db 2041 CAATCTTAACTGTAATCAATCAAGTTACTTGTGTTATCTTCAGAAACATGGGCTGGC 2100
Qy 177 SerTyrAspAsnArgAsnGlyAsnMetAsnHisLysAspProGln----- 190
Db 2101 TCTTACGATTAACAGAAATGGCAATATGATCAATTAACCGCA-GGTATGCTTGAAGTTG 2159
Qy 190 ----- 190
Db 2160 AAAGTATGTCTCTGTTACTAGTATATGATATAGTTACTGTTTACGTTTGTATATA 2219
Qy 191 -----TyrProHisSerTyrGluAspPheLysGlyAspValGlnProSerLysVal 207
Db 2220 TTCTTACAGTATCTCTCATATATGAAGACTTTAAAGAGATGTCCAGCCAAAGTAAAGTT 2279

QY 208 LeuTrrpIleGlyPheProProThrAlaThrgInCyAsnAspGluGlnIleLeuHisAsn 227
DB 2280 CTGGGATTTGGGTTCCCTCCCTACCTGCTACCAATGCAATATGAGCAAAATTCGCACAAT 2339
QY 228 AlaMetIleLeuPheGlyGluIleGluArgValIleSerTyProSerArgAsnPheAla 247
DB 2340 GCGATGATCTCTTTGGTGAAGATGAGAGGGTAAAGTTACCATTCAAGAAATTTTGCA 2399
QY 248 LeuValGluPheArgSerAlaGluGluAlaArgGlnCysIleGluGlnIleArg 267
DB 2400 CTGTGGAGTTTGAAGACGGGAGGAGAGCTGCCAAATGCAAGAGAGCCCTTACAGGGAGAG 2459
QY 268 LeuPheAsnAsnProArgIleValIleMetTySerAsnAspGluLeuProProGluGln 287
DB 2460 TTATTCAATATCTAGAAATCAAAATTAATGACTCAAAACATGATGCTCCCTGAGCA 2519
QY 288 AspAspThrSerPheTySerGlyMetIleArgSerArgThrAspMetPheAsnAsnAsp 307
DB 2520 GACGATCTAGTTTCTCTGCTGATGAACCGTCAAGACAGATATGTTCAATTAATGAT 2579
QY 308 ProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArgProLeuArg 327
DB 2580 CCTTCATTTGATCTCTCTCATTTCTAATGAAATTCCTGGGTCTATGAGGCCCTCAGA 2639
QY 328 GlyThrAsnGluArgSerTyraAsnGlyAlaGluTyraAsnAspValAlaGlyArgPro 347
DB 2640 GGATCAAAATAGCGTTATATTAATGATGAGAAATACATACGTTGTGTAGAGAACCA 2699
QY 348 AsnTrrArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGly 367
DB 2700 AACTGAGAGAGGCACTGCAATAGAAAGCTGAATATCTCCATCTCCAAACGACCTGGA 2759
QY 368 IleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTyr 387
DB 2760 ATCTCCCATCTCTCGTCAAGAGTACGAGCGCCCTATGAGTCAAAACCCGATTTCTGG 2819
QY 388 GlyGlyTyraAspProAlaGlnLeuValArgGluSerTyraArgThrArgAspArgIleSer 407
DB 2820 GAAGGATATATCTGCTCAGTGTGTGAGAAAGTAAACGAAACCAAGAAAGATGATCA 2879
QY 408 ValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAla 427
DB 2880 GTGACCGTTTACTCCATAGGAGTGTGATGAGAGGTCATTTGGTCAAGGTTCACTTGTCT 2939
QY 428 AlaArgProIleArgGlyProProAspSerAspHisIleTrrArgGlyMetIleAlaIle 447
DB 2940 GCTAGACCTATCCGTGGCCCCCTGATTCGATCATATGAGAGGAATGATTGCCAAG 2999
QY 448 GlyGlyThrProValCysGlyAlaArgCysValProMetGlyIleGluThrIleArg 467
DB 3000 GGTGGAATCTCCCGTGTGTGTCTGTGTGATCTATGGGAAAGGGGATTTGAAACCTAA 3059
QY 468 Leu----- 468
DB 3060 CTGTGAGTACTAATTTCTAGACATTTAACCTTCTAGTGTTCCTTTTACAGACCATTT 3119
QY 468 ----- 468
DB 3120 ATATATTTCCATTTCAATTCGATGAGAGTAACATTAATATAGTAGTACATTTTATTT 3179
QY 468 ----- 468
DB 3180 TTACTATTAATCTTTAGTTTCTGAGATGCTTGAATTTTCAATGATGATTTTCAATTTT 3239
QY 469 -----ProGluVa 471
DB 3240 GCATGGCCCTCAATTAATGACTTTTCTTTTAAATATGATTTATAGCCCTGAGGT 3299
QY 471 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaIleValIleValAlaIle 491
DB 3300 CGTCAATTTGTTGAGCAAGAACTGATTTGATATGCTCGTAAACATTAACCCCGTTCAT 3359

QY 491 eGlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrrhrgI 511
DB 3360 TGGATGTAGATCTGTTTTCCTTACCAACAGGAAAGATTTTGGCTTTACCTGA 3419
QY 511 uPheLeuArgTyraLeuSerSerTyraAspArgAlaGlyValAlaIleValAspAspGlyTh 531
DB 3420 ATTTCTCCGGTACCTTAGCTCAAAAGATCGGGGGGTGTGTCCAAATTAATGATGATGTCAG 3479
QY 531 rThrLeuPheLeuValProProSerAspPheLeuThrAspValLeuGlnValThrArg 551
DB 3480 AACTTTATCTTGGTCCCTCCATCAGATTTCTTAATCTGATGATCTCAAGTGCACCGTCA 3539
QY 551 nGluArgLeuTyraGlyValValIleuLeuLeuProProAlaValProValThrAlaSer 571
DB 3540 AGAACGCTATATGAGTGTGTGTCTCAAGTTTACCCCGCAAGCCGTTCTGTTCAGCATC 3599
QY 571 rTyraArgGlnGluSerGlnSerAsnProLeuHisTyraMetAspGlnAlaArgAspSer 591
DB 3600 ATACAGACAAAGATCTCAATCCATCTCGATTAATAGATCAAGCCCGGATTCACC 3659
QY 591 oAlaAsnAlaSerHisSerLeuTyraProArgGluAsnTyraArgGlyAlaProGly 611
DB 3660 TGGCAATGCTAGTACACAGTTTATCTCTTGGGAAATTCATTAGGGGTGCACAG 3719
QY 611 uHisLeuThrAlaAlaSerTyraSerProSerValSerGluProLeuArgIleProAsnAsn 631
DB 3720 ACATTTGACAGCTGCTTCAAAACATCTGTACCGAGCTTCCAGAAATCTTAATATGC 3779
QY 631 aAlaProGlnAlaGlyValSerLeuThrProGluLeuAlaThrLeuAlaSerIleLe 651
DB 3780 AGCGCTCAAGCGTGGGTTATGTTTATCTCGAGCTTTTATAGCATCTGGCATTTCT 3839
QY 651 uProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetSerGlyProSerTh 671
DB 3840 CCTCGCAACTTCTCAACCTGCTGCCCTGAGAGTCAACCACTATCTCAGACCTTCAAC 3899
QY 671 rValAlaSerThrAlaHisGlnSerAsnGlyLeuTyraAsnGlyAlaProSerGlnAl 691
DB 3900 AGTTGTTTCCACGACATCTGATGCAATGACCTGATCAATGAGAAAGCAACGCTCAAG 3959
QY 691 aTrrPysArgGlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnTyrg 711
DB 3960 TTGGAAGAGGTCACAAACAGTTTCAATGATGCTCAATTCAGTCAATCAACAAATACG 4019
QY 711 yAsnGlnTyrrhProAlaGlyGlnLeuProProProProSerArgTyrrProProAlaSe 731
DB 4020 AATTCAGTACACTCCAGCTGGGCACTACTCTCTCTGCGTTACCTCCACTTC 4079
QY 731 rAsnAsnProAsnTyrrhSerGlyMetValHisGlyAsnMetGlnTyrrGlnSerGlnSe 751
DB 4080 AAACAAACCCCACTACATGAGTAATGATGCTGCAACATGCAATACCAAGACCAATC 4139
QY 751 rValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyrrSerMetTy 771
DB 4140 TGTTAACAAGCCTCAGCTGTCTCCGTACCAATATGCTCATTAATTAATTTCCATGTA 4199
QY 771 rThrGlnGlySerSerAsnHisProValSerGlnProMetValGlnGlnTyrrGlnProG 791
DB 4200 CACTCAGGGTTCGTCAAATCATCTGTTCCTGAGCCCAAGTGCAGCAATACCAACCA 4259
QY 791 uAlaSerMetProAsnGlnAsnTyrrGlyProIleProSerTyrrGlnGlnAlaAsnPheH 811
DB 4260 AGGTCATAGCCAAACCAAACTATGATGCTCAATTCCAATGATTCAGCAAGCTTAATTTCA 4319
QY 811 eGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetG 831
DB 4320 TGGCGTAACAACAATCAGGACAGAACTTAACCTTCCAAATTTCAAGCTGCCATGCA 4379
QY 831 nProProAlaAspGlyValAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMe 851
DB 4380 ACCACCAAGAGATTAAGCAAAATTTAGAGCCAAACCAAGCACTACGATTCAGACCTAT 4439
QY 851 tIleSerGlyAspGlyGlnGlyThrThrAspGlyGluValAspIleAsnGlnArgTyrrG 871

Db 4440 GATCTCTGGGATGTCAGGGTACACAGATGGGAGTGCATAGAAATCAGAGATACCA 4499
 Qy 871 nserThleuGlnpheIaIaIaenleuLeuGlnIleGlnGlnIleGlnGlnI 891
 Db 4500 GTCAACACTACAAATTTGCAGCAACCTCTTCTCCAGATACAGCAAGAACAGCAACA 4559
 Qy 891 nserSerGlyThrProIaGlyGlnGlyPro 901
 Db 4560 GTCTTCAGGTACTCCGCTGACACAGGGCCT 4590
 RESULT 4
 ABA95168/c
 ID ABA95168 standard; DNA; 801 BP.
 AC ABA95168;
 XX
 DT 20-MAY-2002 (first entry)
 DE Arabidopsis floral induction gene (FPA) antisense fragment.
 XX
 KM FPA; floral induction; photoperiod; plant; flowering; FLC; FRI; frigida;
 KW flowering locus C; transgenic; gene; antisense; ds.
 OS Arabidopsis thaliana.
 XX
 OS WO200212518-A2.
 PN 14-FEB-2002.
 PD
 XX
 PF 02-AUG-2001; 2001MO-US024427.
 XX
 PR 03-AUG-2000; 2000US-0222550P.
 XX
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Amasino RM, Schomburg FM, Michaels SD, Patton D;
 DR WPI; 2002-227160/28.
 XX
 PT Novel isolated DNA sequence comprising coding sequence for floral
 induction gene (FPA) gene which controls flowering time in plants, useful
 for altering flowering time of plants and for downregulating flowering
 locus C (FLC) mRNA activity.
 PT
 XX
 PS Claim 5; Page 37; 39pp; English.
 XX
 CC The invention relates to the FPA gene (floral induction promoter in
 CC plants during both long and short day photoperiods) from Arabidopsis
 CC thaliana. The FPA coding sequence is useful for altering flowering time
 CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
 CC activity. The FPA gene can accelerate flowering in several late-flowering
 CC mutant backgrounds and can fully compensate for addition of two naturally
 CC occurring genes, frigida (FRI) and FLC which confer late-flowering
 CC phenotypes. Over expression of FPA can compensate for the delaying effect
 CC caused by short days on floral induction, and decreases FLC mRNA in
 CC plants containing the FLC gene. Fragments of the FPA gene can also act to
 CC decrease activity of an endogenous FPA gene by modifying the expression
 CC of the endogenous FPA gene and expression of a portion of polypeptide
 CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
 CC present sequence represents the A. thaliana FPA gene antisense fragment
 XX
 SQ Sequence 801 BP; 206 A; 207 C; 166 G; 222 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7.51e-68 Length: 801
 Score: 1436.00 Matches: 265
 Percent Similarity: 99.62% Conservative: 0
 Best Local Similarity: 99.62% Mismatches: 1
 Query Match: 29.91% Indels: 0
 DB: 6 Gaps: 0

US-09-920-705-3 (1-901) x ABA95168 (1-801)

Qy 197 AaspheIySgIyAspValGlnProSerIySValIleuTrpIleGlyPheProProThAla 216
 Db 799 GACTTTAAAGAGAGATGTCAGCCAAAGTAAAGTTCTGTGATGGGTCTCTCTACTGCT 740
 Qy 217 ThrGlnCyAsnAspGlnIleuGlnIleuHISaenIaMetIleuPheGlyGlnIleGln 236
 Db 739 AAcNAATGCAATGATGAGCAATTTTGCACAAATGCGATGATATCTTTGTGAGATCGAG 680
 Qy 237 ArgValIlySerIyProSerIyArgAsnIleuValGlnIleuPheSerIyAspIleuGln 256
 Db 679 AGGTAAAAAGTTACCCATCAAGAAATTTTGCATCTTGAGTTGAGATTAGAGCGCGAGAA 620
 Qy 257 AlaArgGlnCyIySgIyGlnIleuGlnIleuGlnIleuPheAsnAsnProArgIleuIle 276
 Db 619 GCTCGCCAAATGCAAGAAAGCCCTACAGGGAGATTATTCAAATATCTTCAGATCAAAATT 560
 Qy 277 MetIySerAsnAspGlnIleuProProGlnIleuAspAspThrSerPheIySerGlyMet 296
 Db 559 ATGTACTCAAAAGATGAGTTGCTCTCTGACAAAGCAATACTAGTTTCTCTGATG 500
 Qy 297 IyAspIySerIyArgThrAspMetPheAsnAsnAspProSerIySValIySerIyProIySer 316
 Db 499 AAACGTCACAGACAGATATGTTCAATATGATCTTCATTGTATCTTCTCTCATCTCT 440
 Qy 317 ThrGlyIleProGlySerMetIyArgProIleuArgIyThrAsnGlnIyAspIySerIy 336
 Db 439 ACTGAAATTCCTGGGTCTATAGAGCCCTCAGAGGTCAAAATGAGCTTCATATATAGT 380
 Qy 337 AlaGlnIyAsnAspValIleGlyIySgIyProAsnIyProIyAspIyProIyAspIy 356
 Db 379 GCAGATATCAATGACGTTGTTGTTAGAGAGCCAAATCGAGAGAGCATCTGCAATGGA 320
 Qy 357 ThrGlyIleuProSerProThrGlyProGlyIleuProSerProAlaGlnIyThr 376
 Db 319 ACTGAAATATCCCATCTCCAAAGAGACTGGAATCTCTCCATCTCTGCAAGATGACG 260
 Qy 377 ArgArgProMetIySerIyAspAsnProAspSerIyProGlnIyIyAspProAlaGlnIleuVal 396
 Db 259 AGGCCCTATGAGGTCAAAAGCCGATTTCTGGAGAGATATGATCTCTGCTCACTGATC 200
 Qy 397 ArgGlnIySerIyAspIyArgIyAspIySerIyAspIyProIyPheThrProMetIyVal 416
 Db 199 AGAGAAAGTAAACCAACAGAGAGATGATCACTGACGGTTTACTCCAAATGGGTGTC 140
 Qy 417 AaspGlnIySerIyPheGlyIyArgIySgIySValIleuIleuArgProIleuArgIyProProAsp 436
 Db 139 GATGAGAGTATTTGTCAGAGTTGCTGCTGACTAGACCTATCCGTGCCCCCTGAT 80
 Qy 437 SerAspHisIleIyPArgGlyMetIleuAlaIySgIyGlyThrProValIySgIyAlaArg 456
 Db 79 TCTATACAAATGATGAGAGATATTTGCCAAGGTGGAATCTCCGCTCTGTGTGCTCTGT 20
 Qy 457 CysValProMetGlyIySgIySValIleuIleuArgProIleuArgProProAsp 462
 Db 19 TGTGTACTTATGGGAAG 2
 RESULT 5
 ABA95170
 ID ABA95170 standard; DNA; 3715 BP.
 AC ABA95170;
 XX
 DT 20-MAY-2002 (first entry)
 DE Arabidopsis floral induction gene (FPA) fragment.
 XX
 KM FPA; floral induction; photoperiod; plant; flowering; FLC; FRI; frigida;
 KW flowering locus C; transgenic; gene; ds.
 OS Arabidopsis thaliana.
 XX
 PN WO200212518-A2.

XX 14-FEB-2002.
 PD 02-AUG-2001; 2001MO-US024427.
 XX 03-AUG-2000; 2000US-0222550P.
 PR
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 PA
 XX Amasino RM, Schomburg FM, Michaels SD, Patton D;
 PI WPI; 2002-227160/28.
 DR
 XX Novel isolated DNA sequence comprising coding sequence for floral
 PT induction gene (FPA) gene which controls flowering time in plants, useful
 PT for altering flowering time of plants and for downregulating flowering
 PT locus C (FLC) mRNA activity.
 XX
 PS Claim 5; Page 38; 39pp; English.
 XX
 CC The invention relates to the FPA gene (floral induction promoter in
 CC plants during both long and short day photoperiods) from Arabidopsis
 CC thaliana. The FPA coding sequence is useful for altering flowering time
 CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
 CC activity. The FPA gene can accelerate flowering in several late-flowering
 CC mutant backgrounds and can fully compensate for addition of two naturally
 CC occurring genes, *frigida* (FRI) and *FLC* which confer late-flowering
 CC phenotypes. Over expression of FPA can compensate for the delaying effect
 CC caused by short days on floral induction, and decreases FLC mRNA in
 CC plants containing the FLC gene. Fragments of the FPA gene can also act to
 CC decrease activity of an endogenous FPA gene by modifying the expression
 CC of the endogenous FPA gene and expression of a portion of polypeptide
 CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
 CC present sequence represents the A. thaliana FPA gene fragment comprising
 CC the FPA promoter and intron
 CC
 SQ Sequence 3715 BP; 1055 A; 728 C; 711 G; 1221 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.97e-25 Length: 3715
 Score: 651.50 Matches: 171
 Percent Similarity: 27.23% Conservative: 0
 Best Local Similarity: 27.23% Mismatches: 0
 Query Match: 13.57% Indels: 457
 DB: 6 Gaps: 1
 US-09-920-705-3 (1-901) x ABA95170 (1-3715)
 QY 1 MetAlaIeuSerMetLysPProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
 DB 1832 ATGGCGTTATCTTGAAGCCATTCAGAGCGATGATTCGGGTTCCAGTAAACAACTCT 1891
 QY 21 TTPValGlySerLeuThrProGluThrThrGlnuSerAspLeuThrGluLeuPheGlyArg 40
 DB 1892 TGGGTGGGTAGCCTTAAGCGCGAGAGAGAGAGATCTGACCGAGTTGTTGGAGA 1951
 QY 41 TyrGlyAspIleAspArgGlyIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
 DB 1952 TAGCGGATATTTGATGAAATCAGCGTGTATCTTCAAGAGCTTTGGGTTATATATCTAC 2011
 QY 61 ArgHisValGlnGluIuValAlaValAlaValGlnGluAlaGlnGlyIleAsnLeuAsnGly 80
 DB 2012 AGACATGTGGAG 2071
 QY 81 SerGlnIleLysIleGlyTyrAlaArgPro----- 90
 DB 2072 AGTCAATTAAGATCGAATCGACACGCGTTGTTCTATCATATCTTCGTTGTTGTC 2131
 QY 90 ----- 90
 DB 2132 TCTAATTTGATGTTCTTTTGTCAAGCATTAATACTTTTGGGAATTCATAGTCAGGT 2191
 QY 90 ----- 90

DB 2192 TCACAAACTTTGATGATGCTGTTTATGTCAAAAATTTCTTGATCTGTTTTTTT 2251
 QY 90 ----- 90
 DB 2252 TTCCATGATGTAACAATCAAGTCGAACCTAGTTTTTTCTATATATAGTCGTTAGCTT 2311
 QY 90 ----- 90
 DB 2312 AAGGCAAACTGATCCGATCGAAACGCTCTTTCTCAAAATTAATTGTTATATCGAACT 2371
 QY 90 ----- 90
 DB 2372 CGGCAAAAGCCAAACACAGAGAAGCTTCGAAAATTTGATGTTAAAGCATATATTAAGTC 2431
 QY 90 ----- 90
 DB 2432 TTAGCAATGAGCTCTCAAAAGAAATACATCAAAACATGTTTACTCTGTTATGCGAA 2491
 QY 90 ----- 90
 DB 2492 GAAAGTTTAATCGATTTGCTTTATCTGCAAACTTGATCCGCGTTTAGCACTGAT 2551
 QY 90 ----- 90
 DB 2552 TCAGTTGCTTCAGATGTTAAATCTCAGACGTTGATGATGATTTGGTACTCCA 2611
 QY 90 ----- 90
 DB 2612 CTCGAATTCGCAATGTGTAATAATTTGAAGCGAGCAACTATCATCGGCCAAACAG 2671
 QY 90 ----- 90
 DB 2672 TAAGAAATTTGAGATTAAAGTTCAAGTTCTCTGCAAAAATCAAAACCGGTGAGAAAT 2731
 QY 90 ----- 90
 DB 2732 TTGCTATGCGGTAGATGATATCAATACTGCACTGCGAAACGAAATTTCTGGCAA 2791
 QY 90 ----- 90
 DB 2792 CCGTATGTCCTTAATCTAGTATCGCTGCGACATATCATCAAGTGTGAGGAGTTT 2851
 QY 90 ----- 90
 DB 2852 GGGCTTTGTTGGCTCGATGCTTCAGAAAGACAAATTAAGTGTGTTAAGCGGCTAAC 2911
 QY 90 ----- 90
 DB 2912 TCTACCAATCAGAAACGCTTATTTGCAAGAACCATGTTGTTCTCAATTCCTCCATCCCTA 2971
 QY 90 ----- 90
 DB 2972 CGTACATCTGGGCTTTCCCATTTAGTCTCTTAGAAGTTGACTCTTCACAAACATTC 3031
 QY 90 ----- 90
 DB 3032 TGGGATGTTGACAGTTGACGTGAGCAAGACAAATATTCATTTGACGAGCACTCAAAATTT 3091
 QY 90 ----- 90
 DB 3092 CAATGCTTCCGGGCTTTAGTACTTAATGAGAGTTCAGGAGTTGTTATTTGGGAC 3151
 QY 90 ----- 90
 DB 3152 TCAGCAATGATTAAGACAGAAATGTTTTCAGAAAGTATTAATGTTCTTCGCGCTTG 3211
 QY 90 ----- 90
 DB 3212 GGGACTACACTTCTCTCTACAGACAAATCTGATGTTTAGGCAAGAACTATTTACA 3271
 QY 90 ----- 90

Db 3272 CTGATTATAGAGAAAGAGAGATGTTCTCTGGGCAATTTATTTGTTAGAA 3331
 Qy 90 ----- 90
 Db 3332 AGCAATTGATGTAATGGTGTCTAGTGTGAAATTAAGTAGTTGTGTCTT 3391
 Qy 90 ----- 90
 Db 3392 AGTTCTTATGTTGATGTTATCTTGTGCAACCTATCTGGTAGTATCGCTTCTT 3451
 Qy 91 ----- AlalyProCyalySserLeuTPValGlyIleGly 103
 Db 3452 ATGCACCTTTCTTGTTCAGAGCAAAACCTTGTAAGGTCTAGGCGGTGGAATCGC 3511
 Qy 104 ProAsnValSerIysAspAspLeuGluGluGluPheSerIysPheGlyIleGlyAsp 123
 Db 3512 CCTATGTCTCCAGAGATGACCTGAGAGAGAGTTGCAAGATTTGGGAAAATCGAGAT 3571
 Qy 124 PheArgPheLeuArgIleArgIleThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
 Db 3572 TTTAGGTTTCTCAGAGAACCAAGACAGCTTTCATTGATTATAGATGATGATGCT 3631
 Qy 144 LeuGlnAlaIysSerMetLeuGlyIlePheMetGlyIleSerPheLeuArgValAspPhe 163
 Db 3632 TTACAGGCTAAGAGCAATGAATGAAAGCCATAGGTTGTTGCGTGTGATTTT 3691
 Qy 164 LeuArgSerGlnAlaPheProIysIys 171
 Db 3692 CTCGGTCCACAGGCCCAAAAAA 3715
 RESULT 6
 ID ABA95169 standard; DNA; 349 BP.
 AC ABA95169;
 XX
 DT 20-MAY-2002 (first entry)
 DE Arabidopsis floral induction gene (FPA) coding fragment.
 XX
 KM FPA; floral induction; photoperiod; plant; flowering; FLC; FRI; frigida;
 KM flowering locus C; transgenic; gene; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200212518-A2.
 PD 14-FEB-2002.
 PF 02-AUG-2001; 2001WO-US024427.
 PR 03-AUG-2000; 2000US-0222550P.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Amasino RM, Schomburg FM, Michaels SD, Patton D;
 DR WPI; 2002-227160/28.
 XX
 PT Novel isolated DNA sequence comprising coding sequence for floral
 induction gene (FPA) gene which controls flowering time in plants, useful
 for altering flowering time of plants and for downregulating flowering
 locus C (FLC) mRNA activity.
 PT
 XX
 PS Claim 5; Page 37; 39pp; English.
 CC The invention relates to the FPA gene (floral induction promoter in
 CC plants during both long and short day photoperiods) from Arabidopsis
 CC thaliana. The FPA coding sequence is useful for altering flowering time
 CC in a plant, and for down-regulating flowering locus C (FLC) mRNA
 CC activity. The FPA gene can accelerate flowering in several late-flowering
 CC mutant backgrounds and can fully compensate for addition of two naturally
 CC occurring genes, *frigida* (FRI) and *FLC* which confer late-flowering

CC phenotypes. Over expression of FPA can compensate for the delaying effect
 CC caused by short days on floral induction, and decreases FLC mRNA in
 CC plants containing the FLC gene. Fragments of the FPA gene can also act to
 CC decrease activity of an endogenous FPA gene by modifying the expression
 CC of the endogenous FPA gene and expression of a portion of polypeptide
 CC encoded by FPA gene can also lead to a delay of flowering in a plant. The
 CC present sequence represents the A. thaliana FPA gene coding fragment
 XX
 SQ Sequence 349 BP; 92 A; 68 C; 81 G; 108 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,12e-16 Length: 349
 Score: 466.00 Conserved: 92
 Percent Similarity: 94.95% Mismatches: 2
 Best Local Similarity: 92.93% Indels: 3
 Query Match: 9.71% Gaps: 1
 DB: 6
 US-09-920-705-3 (1-901) x ABA95169 (1-349)
 Qy 1 MetAlaLeuSerMetIysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
 Db 1 ATGGCGTTATCTATGAAACCAATTCAGAGCCGATGATTCGGTTCCAGTCAAAACATCTT 60
 Qy 21 TTPValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
 Db 61 TGGGTGGTAGCTTAACCCGAGAGCGACAGATGATGATGACGAGTGTGTTGGAAGA 120
 Qy 41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyr 60
 Db 121 TACGCGATATTGATGATGAAATCAGCGGTATCTTCACAGGCTTTCGTTTATATATCTAC 180
 Qy 61 ArgHisValGluGluAlaValAlaAlaIysGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
 Db 181 AGACATGTGGAGAGAGCGAGTGCAGCCAAAGAGCTCTTCAGAGAGCAAAATTTGAATGA 240
 Qy 81 SerGlnIleIysIleGluTyrAlaArgProAlaIysProCysIysSerLeuTPVal 99
 Db 241 AGTCAAATTAAAGATGAAATACGACGACCGGTT-----TGTCTTATCTATATCTT 291
 RESULT 7
 ID ABZ21692 standard; CDNA; 6836 BP.
 AC ABZ21692;
 XX
 DT 27-FEB-2003 (first entry)
 DE Human RBM15-MKL1 fusion protein encoding cDNA SEQ ID NO.1.
 XX
 KM Human; RBM15; RNA binding motif protein 15; megakaryoblastic leukaemia 1;
 KM MKL1; fusion protein; acute megakaryoblastic leukaemia; AMKL; cytostatic;
 KM c(1; 22) chromosomal rearrangement; gene therapy; chromosome 1p13;
 XX
 OS Homo sapiens.
 XX
 PI Key Location/Qualifiers
 FT CDS 84..5735
 FT /*tag= a
 FT /product= "RBM15-MKL1 fusion protein"
 XX
 PN WO200288309-A2.
 PD 07-NOV-2002.
 XX
 XX 23-APR-2002; 2002WO-US012797.
 XX 27-APR-2001; 2001US-0286910P.
 PR (SJD)- ST JUDE CHILDREN'S RES HOSPITAL.
 PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
 XX

PI Morris SW, Ma Z, Hitzler JK;
 XX WPI: 2003-103455/09.
 DR P-PSDB; ABP56065.
 PT New RNA-binding motif protein-15 (RBM15)-megakaryoblastic leukemia-1
 PT (MKL1), MKL1-RBM15-S and MKL1-RBM15-S+AE fusion protein, useful for
 PT identifying agents useful for treating patients with acute
 PT megakaryoblastic leukemia.
 XX
 PS Claim 2; Page 53-62; 109pp; English.
 XX
 CC The present invention describes an RNA-binding motif protein-15 (RBM15)-
 CC megakaryoblastic leukemia-1 (MKL1) fusion protein, a MKL1-RBM15-S fusion
 CC protein, and a MKL1-RBM15-S+AE fusion protein associated with acute
 CC megakaryoblastic leukemia (AMKL). Also described: (1) an antibody that
 CC specifically binds to the RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE
 CC fusion proteins; (2) a non-human transgenic animal that has been altered
 CC to express a gene encoding a RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE
 CC fusion protein; (3) identifying an agent capable of binding to a RBM15-
 CC MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE fusion protein; (4) detecting the
 CC t(1;22) chromosomal rearrangement associated with AMKL; and (5) screening
 CC for agents capable of (selectively) inhibiting the activity of a fusion
 CC protein arising from the t(1;22) chromosomal rearrangement associated
 CC with AMKL. The fusion proteins have cytoskeletal activity and can be used
 CC in gene therapy. The RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AE fusion
 CC proteins and nucleotide molecules are useful for designing and preparing
 CC agents that specifically inhibit the expression of the RBM15-MKL1 or MKL1-
 CC RBM15 genes in cells for therapeutic and other purposes. The transgenic
 CC animals are useful for identifying and testing carcinogenic or
 CC therapeutic compositions. The methods are also useful for detecting the
 CC t(1;22) chromosomal rearrangement associated with AMKL, or for
 CC identifying agents useful for treating patients with AMKL. The antibodies
 CC can be used to selectively kill cells expressing RBM15-MKL1, MKL1-RBM15-
 CC S, or MKL1-RBM15-S+AE fusion protein. RBM15 is located to chromosome
 CC 1p13, and MKL1 is located to chromosome 22q13. The present sequence
 CC encodes the RBM15-MKL1 fusion protein of the present invention
 XX
 XX Sequence 6836 BF; 1473 A; 2089 C; 1963 G; 1311 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 2,11e-08 Length: 6836
 Score: 336.50 Matches: 237
 Percent Similarity: 34.37% Conservative: 127
 Best Local Similarity: 22.38% Mismatches: 371
 Query Match: 7.01% Gaps: 324
 DB: 8 Indels: 48

US-09-920-705-3 (1-901) x AB21692 (1-6836)
 QY 7 ProPheArg-----AlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
 DB 1152 CCTTTCGAGAGAGTGATGAGATTTTCCACCGAGATATCAGACGTAACCGGACGCTC 1211
 QY 21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
 DB 1212 TTCTTGGAACCTCAGACATCACTGTAACGAGAGTATTGAAGAGGCGCTTGATGCGC 1271
 QY 41 TyrGlyAspIleAspArgIleThrVal---TyrSerSerArgGly-----Phe 55
 DB 1272 TTGGAGCTATCAGAGAGTATGATCAAGAGCGCTTCGCGGCGAGACTAGTACTTAC 1331
 QY 56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaValGluAlaLeuGlnGly 75
 DB 1332 GCGTTCTCAATTTGAGAACTTGAATATGCTCACCGGCGCAAAATTAGCAATGTGCGC 1391
 QY 76 AlaAsnLeuAsnGlnSerGlnIleValIleGluTyrAlaArgProAlaLysProCysLys 95
 DB 1392 AAAATTATATTCGGATTCCTATCAAAATGTTATGTAA---GCTAACACCCACACACC 1448
 QY 96 SerLeuTrpValGlyGlyIleGlyProAsnValSerLysAspAspLeuGlnGluPhe 115
 DB 1449 CGCCTCTGGGTGGAGGCGTGGGACCTTGGTTCCTGCTGCGCCGAGACAGAAATTT 1508
 QY 116 SerLysPheGlyLysIleGluAspPheAsnArgGluArgLysThrAlaPheIle 135
 DB 1509 GATGATTGGTGCACCATACGACCATATGACTACGAAAGGTAGTATGGGACTATATTC 1568
 QY 136 AspTyrTyrGlyLeuTyrAspAspAlaLeuGlnAla---LysSerMetAsnGlyLysPromet 154
 DB 1569 CATTATGAAAGCCTTGATGATGAGGATGCTGCTGCTGAGACCATATGGGGCTTCCACTT 1628
 QY 155 GlyGly-----SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysValGlu 172
 DB 1629 GGTGCCCCAGATGACGACGCTTAGATGACTTGGCCGACACGAAATCGTTACACAGAG 1688
 QY 173 GlnTrp-----AlaGlySerTyrAspAsn 180
 DB 1689 CAGTATCTGCAGACCTCTGCTGCTGACTTATGACCTGTGACAGATGCTTTTGGACAT 1748
 QY 181 Arg-----AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyr 195
 DB 1749 CGGGACACAGACCTTTGAGGGGTGCTGGGATGAGACACA-----CCCTTACTATAC 1802
 QY 196 GluAspPheLysGlyAspValGlnProSerLysValLeuTrpIleGlyPheProThr 215
 DB 1803 AGAGATCGTATAGGAGACCTTATCTGACTGTAT---TGGGTGCACCCACCCACCCCA 1859
 QY 216 AlaThrGlnCysAsnAspGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIle 235
 DB 1860 GTCCGA-----GAGCCAGACACTCGGACGTGACCTTCTGCTGCTGTAT 1907
 QY 236 GluArgValLysSerTyrProSerArg-----AsnPheAlaLeuValGluPheArgSer 253
 DB 1908 GAGCCATGATATGCTTAATATGATGACGAGGATGTGTGCTTTCGACCGGACAGAGT 1967
 QY 254 AlaGluAlaAlaArgLysCysLysGlyGlyLeuGlnGlyArgLeuPheAsnProArg 273
 DB 1968 GATGAGATCTGCGCCAGACAGACAGAC-----CAGCCTAAG 2003
 QY 274 IleValIleMetLysSerAsnAspGluLeuProProGlnGluAsp-----Asp 289
 DB 2004 AACCGA-----AGCTGCTCGAGAGAGTGAAGTGAAGACATCATCTGAT 2045
 QY 290 ThrSerPheTyrSerGlyLysMetLysArgSerArgThrAspMetPheAsnAsnAspProSer 309
 DB 2046 AGCTCTCTGAGAGTATGACCGCCACGAAACG-----CAC 2081
 QY 310 CysValSerSerPro---HisSerThrGlyIleProGlySerMetArgProLeuArgGly 328
 DB 2082 TGGCTCTCTCTCTGACCGGACGTCAGAAATGAGCAGTACCGGATGCTTACACAGAG 2141
 QY 329 ThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValAlaGlyLysGluProAsn 348
 DB 2142 GACATATGATCGATCT-----TCCCGCTCTCTCTTGAAGAGCCCTCTC 2183
 QY 349 TrpArgArgProSerLysAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIle 368
 DB 2184 -----CAATCAAGACGAGCAGAGTATTT----- 2210
 QY 369 LeuProSerProAlaGlnGlyThrArgArgPrometArgSerAsnProAspSerTrpGlu 388
 DB 2211 -----GAGAAAGACGAGGTGACAGCAGAC---CGTAAAACTCTGATAGCTGAA 2261
 QY 389 GlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAspGlySerVal 408
 DB 2262 -----CGAGATAGAGACACCGGACCAACTGCTCCACT 2294
 QY 409 AspGlyPheThrPrometGlyValAspGluArgSerPheGlyArgGly-----Ser 425
 DB 2295 GAGGGAAGAAAGCCTCTGAAAAAGAACCGCTCTGATGAGAGTGCACCTAGCACACAC 2354
 QY 426 ValAlaAlaArgProIleArgLysPro----- 434
 DB 2355 ACTGCTCTCTCAAGCTGAAGTCCCGTCCAGAAACAGATGGGGGACAGCCCTGTG 2414

444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

D	b		 3429 CTTCCGATGGCGCGGAGATTCCAGAAATGCTTTTCTTCGGACAGCAGCCTCTCTCCT	 3488
O	y		722 ProProPProSerArgTyProProAlaSerAsn-AsnProAsnTyThrSergIyMetVa	741
D	b		3489 CCCCACTCTGTGCTGCCCTCCAGCCTCACCAATGAACA-----	3528
O	y		741 IHisGIyAnMeGIntYrGIInSerGIInSerValAsnMeCProlInLeuSerProleuPr	761
D	b		3529 -----CTATCCCCACTGCC	3542
O	y		761 oASnMePProHIsAenAntYrSerMeTYrThrGIInGLysSerSerAsnHisProValSe	781
D	b		3543 AAGTCCACC-----CCACACTCATTAAG	3566
O	y		781 r-GlnProMetValGIInGLInTYrGIInProGIInAlaSerMeCProAsnGLInAsnTYrGLYP	801
D	b		3567 CAAGAACA---ACCCAAGCTGCACGTAGAAAGTCACAGCGCAGCAAGAAAGCCAGAGG	3623
O	y		801 ro-IIEProSeryTYrGIInGLInAlaAsnPheHisGLyValThrThrAsnGLInAlaGLInAsn	820
D	b		3624 CTGAAGCCAAAGTGAAAGAACCTCAAGTACCAC-----CAGTAC	3662
O	y		821 LeuAsnProSergInPheGLInAlaAlaMeCGLInProProAlaAPLySAIAEnLeuGLu	840
D	b		3663 ATCCCCCGGACACGAAGCAGGAGAGGGGGGACCCCCCATGACTATCTCTACGCCAAG	3722
O	y		841 ---ProGLInAsnGLInAlaLeuArGLeuGLInProMeTIleSergIyAspGLInGLYthr	859
D	b		3723 ATCTGCAGCAGCAGCAGCCTCTCTCTCAG-----	3752
O	y		860 ThrAspGLyGIInValAspLYAsnGLInArGYrGLInSerThrLeuGLInPheAlaIAEn	879
D	b		3752 -----	3752
O	y		880 LeuLeuLeuGLInILeGLInGLYLeuGLInGLInGLIn-----	891
D	b		3753 ----CTGCAGATCTCAACACGACGACAGACAGACACACAACTACAGGCCATCTCG	3806
O	y		892 -----SergGLYthrPro	896
D	b		3807 CCTGCCCGCCAAAGTACAGCAGCGAGCCCTTGGAAGCAGCGGAGCCCCC	3857
RESULT 8				
AAZ51263				
ID	AAZ51263	standard; cDNA, 2190 BP.		
XX	XX			
AC	AAZ51263;			
DT	06-JUN-2000	(first entry)		
DE	Human RNA-associated protein-14 (RNAAP-14) encoding cDNA.			
XX	XX			
KW	RNA-associated protein; RNAAP; human; clone 1748626; cytosolic;			
KW	immunosuppressive; antiinflammatory; keratolytic; neuroprotective;			
KW	antidiabetic; osteocytic; hepatotropic; antiparasitic; anti-HIV;			
KW	antiatherogenic; antirheumatic; antiarthritic; opthalmologic; autoimmune;			
KW	antimicrobial; cell proliferative disorder; inflammation; cirrhosis;			
KW	acneic keratosis; bursitis; arteriosclerosis; artherosclerosis;			
KW	hepatitis; myelofibrosis; primary thrombocythemia; psoriasis; cancer;			
KW	mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease;			
KW	allergy; rheumatoid arthritis; parasitic infection; ss.			
XX	XX			
OS	Homo sapiens.			
XX	XX			
FH	Key	Location/Qualifiers		
FT	CDS	164..1855		
FT	/tag=	a		
FT	/product=	"Human RNA-associated protein-14"		
XT	/note=	"Derived from STOMUT02 library"		
PV	MO200011171-A2.			
XX	XX			

PD 02-MAR-2000.
 XX 20-AUG-1999; 99MO-US019361.
 XX 21-AUG-1998; 98US-0097550P.
 PR 12-JAN-1999; 99US-0115639P.
 XX (INCY-) INCYTE PHARM INC.
 XX Hillman JL, Yue H, Tang YT, Corley NC, Guejler KJ, Gorgone GA,
 PI Patterson C, Baughn MR, Lal P, Bandman O, Reddy R, Azimzai Y,
 PI Shih LT, Yang J, Lu DM;
 DR WPI; 2000-237651/20.
 DR P-PSDB; AAY70233.
 XX
 PT Human RNA-associated proteins useful in diagnosing, treating and
 PT preventing cell proliferative, autoimmune, inflammatory and infectious
 PT disorders.
 XX
 PS Claim 9; Page 116-117, 123pp; English.
 XX
 CC The present sequence is the cDNA encoding human RNA-associated protein-14
 CC (RNAAP-14), identified in Incyte clone 1748626, derived from STOMT02
 CC library. It is expressed in reproductive, nervous, gastrointestinal, and
 CC hematopoietic/immune tissues. It has cytosolic, immunosuppressive,
 CC antiinflammatory, antiarteriosclerotic, hepatocytic, keratolytic,
 CC neuroprotective, antiparasitic, anti-HIV, antiallergic, antirheumatic,
 CC antiviral, antitubercular, ophthalmological and antimicrobial activity.
 CC RNAAP antibodies are useful for diagnosis of diseases associated with
 CC altered expression or activity of RNAAP. It is used to treat cell
 CC proliferative, autoimmune, inflammatory and infectious disorders, like
 CC actinic keratosis, burns, arteriosclerosis, artherosclerosis,
 CC cirrhosis, hepatitis, myelofibrosis, mixed connective tissue disease
 CC (MCTD), psoriasis, primary thrombocytopenia and cancer, HIV, allergies,
 CC rheumatoid arthritis, uveitis, Crohn's disease, and bacterial, viral and
 CC parasitic infections
 XX
 SQ Sequence 2190 BP; 515 A; 616 C; 633 G; 426 T; 0 U; 0 Other;
 XX
 Alignment Scores:
 Pred. No.: 4,22e-07 Length: 2190
 Score: 302.00 Matches: 157
 Percent Similarity: 33.67% Conservative: 78
 Best Local Similarity: 22.49% Mismatches: 232
 Query Match: 6.29% Gaps: 27
 DB: 3
 US-09-920-705-3 (1-901) x AAZ51263 (1-2190)
 QY 11 AspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrThr 30
 DB 170 GAGGATATACCAAGCGGCGGCGCAACCTTCATTCATGTAACCTGGACACACAGCATCT 229
 QY 31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrVal--- 49
 DB 230 GAGGTGAGCTGGCAAGGCGCTTCGAGAAATATGCGCATCAAGAGGAGTGTATCAAG 289
 QY 50 TyrSerSerArgGly-----PheAlaPheIleTyrTyrArgHisValGluGlu 65
 DB 290 AGCGCTGCGGCGGCGGCGGCGGCTATGCTTCCTCAAGTTCCGAACCTGGACATG 349
 QY 66 AlaValAlaAlaValGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleValIle 85
 DB 350 GCCCATATGAGCTTAAGTGGCCATGTCGCGCGGAGTGTGTGTCGAACCCCATTAAGATA 409
 QY 86 GluTyrAlaArgProGlyAlaSerProGlyValSerLeuTyrValGlyGlyIleGlyProAsn 105
 DB 410 GGCTATGGCAAG--GCCAACCCCAACCACTGCTTCTGGGTGGGTGGCTGGAGACTTAAC 466
 QY 106 ValSerTyrAspAspLeuGluGluGluGluPheSerTyrPheGlyLysIleGluAspPheArg 125
 DB 467 ACCTCACTGGCGGCTGTGGCGCCAGAGTTTGAACCGCTTGGAGAGCATTGGACCATTTGAT 526

QY 126 PheLeuArgGluThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
 DB 527 CACGTCAAGAGATAGCTTTGCTTATTCAGTATGAGAGACTTGGACCGACCGCC 586
 QY 146 Ala---LysSerMetAsnGlyLysProMetGlyGly-----SerPheLeuArgValAsp 162
 DB 587 GCTTGCTGCTAAATATGAGGGGTTTCCCTTGGGTGAGACGACCGAGGCTCCCGTGGAT 646
 QY 163 PheLeuArgSerGlnAlaProLysGluGlnTyrPheGlySerTyrAspAsnArgAsn 182
 DB 647 TTGCGCAAGACAGAG----- 661
 QY 183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
 DB 662 -----GAGACTCGGTACCCCGACGACTAC----- 685
 QY 203 GlnProSerTyrValLeuThrPheProProThrAlaThrGlnCysAsnAspGlu 222
 DB 686 CAGCCTTCGCCACTC-----CTGTGCATTTATGAGTCTGCACAGATGCA 730
 QY 223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrPro 242
 DB 731 TACACCGGCAACCGCAAC-----CTGACGCGGACCTGTGGGGAGACAGACGCCCA 784
 QY 243 -----SerArgAsnPheAlaLeuValGluPheArgSerAlaGlu 255
 DB 785 CACCTTCTGTACTAGACCGGACCGGACCTTTTGGAAAGGAGCTGACACGCCCACT 844
 QY 256 GluAlaArgGlnCysLysGluGluGlnGlyArgLeuPheAsnAsnProArgIleLys 275
 DB 845 AAAAGCTGACCGCGGCAACACCTTGAAGGC----- 877
 QY 276 IleMetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGly 295
 DB 878 -----TACAGTCCG 886
 QY 296 MetLysArgSerArgThrAsp--MetPheAsnAsnAspProSerCysValSerSerProH 315
 DB 887 TCAGTCCGACCGGCGGAGGTGTAGCGTGGGGGCGAGATGAGACCGTGTGGCCCAAG 946
 QY 315 sSerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAs 335
 DB 947 C-----CTGGG-----AAAGAGCGCGAAA 967
 QY 335 nglyAlaGluTyrAsnAspValValGlyLysGluPro-----AsnTyrAs 350
 DB 968 CGGAGAAAGCTTTCAGTACCGGTGGAGAGACCAACCATTCACATATGAGAAACGAGT 1027
 QY 350 gatGProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro----- 366
 DB 1028 AGGACCAAGGAGAGTGGGAGAGTCAAGCGGGGCTTCGACCGCACCCCTGAGCGCAGC 1087
 QY 367 -GlyIleLeuProSerProAlaGlnGlyThrArgArg----- 378
 DB 1088 CGCAAGAGAACAACATCCAGTGAAGGACCAAGAGTCCAGACGAACTCCCTCAGAAC 1147
 QY 378 ----- 378
 DB 1148 AGCAGATAGGGGCTGAGAGACGGGGCACACACACACACAGAGGCTGACATCT 1207
 QY 379 -----PrometArgSerAsnPr 384
 DB 1208 TCCACAGGAGAGAGCAAGAGACAGCGCAATCACCGGACAGAGCGGAGCC 1267
 QY 384 AspSerTyrGluGluTyrAspProAla--GlnLeuValArgLysSerLysArgThrArgA 404
 DB 1268 AAGCTCTGAAAGAGCAAAACACAGACCAAAAGCTTAAGAAATTTTCAAGATACGCT 1327
 QY 404 rghspGlySerValAspGlyPheThrPrometGlyValAspGluArg----- 419
 DB 1328 CAGACACTACAGCTGGGTGAGATGAGGCTTCTGTGTGTAAAGAAACAGCTTCCCCACAG 1387

```
QY 420 -----SerPheGlyArgGlySerValAlaIala----- 428
DB 1388 TCTATGCAATATCTCAGAGGGGAGACAGGGGGTGTATCGACGTCTCTCAAGACCACT 1447
QY 429 -----ArgProIleArgGlyPro-ProAsp 436
DB 1448 TCTGGGAGCAAGCTGACCCAGCTGAAGATCGCCAGCGCTTGACACTGACCAAGCCAG 1507
QY 437 SerAspHisIleIleTPrpArgGlyMetIleAlaGlyIleGlyThrPro----- 451
DB 1508 CTTGACAGATCAGACAGCAAGCCATC-----AAGCAGGGGAGCCCAAGCGCTATGCGGTC 1561
QY 452 ValCysSerAlaArgCysValProMetGlyLysGlyIleGlyThrLysLeuProGluVal 471
DB 1562 CTCTTACGACCAAGCAAGCAAGCCAGCGCTTGACACTGAG--GGGATGCCACAGTA 1618
QY 472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIle 491
DB 1619 GAGCCCGGCTCGACAGGCGGCTTCTCAGGAACTG----- 1654
QY 492 GlyCysGluIleValPhePheValProAspArgGluAspPheAlaSerTyrThrGlu 511
DB 1654 ----- 1654
QY 512 PheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAsp----- 528
DB 1655 ---GTCTCTTACTTGAACAGAAAGAGCGGCGGAGGTATCGCTTGCCAGTGGGGGG 1711
QY 529 -----AspGlyThr---ThrLeuPheLeuValProProSerAspPheLeuThr 543
DB 1712 TCCAAAGGAGAAACGAGCAAGCATCTCTACGCTTCCACCTCGACCTTTCACAG 1771
QY 544 AspValLeuGluIleValThr-----ArgGluGluArgLeuTyrGly 556
DB 1772 CAGTACTCTCAGTCAGACACTAAGACATTGGCGCACTGAAAGAAACATGCTGATA 1831
QY 557 ValValLeuLys-----LeuProProProAlaValProValThrAlaSerTyrArg 573
DB 1832 GTCATCGTCAGAGACACTGCTGACCAAGCTGCTTCCAGCGCTCANTGTTGTGTA 1891
QY 574 GlnGluSer-----GlnSerAsnProLeuHisTyr 583
DB 1892 CAAAGCAGATTATTAAATCTGATCCCTCTCTACCTTACCACTTT 1939

RESULT 9
ACN37956
ID ACN37956 standard; cDNA; 3127 BP.
AC
ACN37956;
DB 18-NOV-2004 (first entry)
XX Tumour-associated antigenic target (TAT) cDNA DAA324333, SEQ ID NO:1186.
DE
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX tumour; diagnosis; cell proliferative disorder; breast cancer;
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX central nervous system cancer; bladder cancer; pancreatic cancer;
XX cervical cancer; melanoma; leukaemia; hybridisation probe;
XX chromosome identification; chromosome mapping; gene mapping;
XX gene therapy; cytostatic; gene; ss.
OS Homo sapiens.
XX
XX PN WO2004030615-A2.
XX
XX PD 15-APR-2004.
XX
XX PF 29-SEP-2003; 2003WO-US028547.
XX
XX PR 02-OCT-2002; 2002US-0414971P.
XX
XX PA (GENTH ) GENEINTECH INC.
```

```
XX Wu TD, Zhang Z, Zhou Y;
PI
PI MPI; 2004-347921/32.
DR
XX New tumor-associated antigenic target polypeptides and nucleic acids,
XX useful in preparing a medicament for treating or detecting a
XX proliferative disorder, e.g. Breast, lung, colorectal, ovarian or
XX prostate cancer or tumor.
XX
XX Claim 1, SEQ ID NO 1186; 7273bp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are
XX overexpressed in cancer tissues compared to normal tissues, and may thus
XX serve as effective targets for the diagnosis and treatment of cancer in
XX mammals. The invention also relates to nucleic acid and polypeptide
XX sequences at least 80% identical to the TAT nucleic acids and
XX polypeptides; expression vectors and host cells comprising a TAT nucleic
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a
XX TAT polypeptide; and methods and compositions for the treatment or
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
XX antibodies, antagonists, binding molecules and compositions are useful
XX for diagnosing or treating a cell proliferative disorder associated with
XX increased TAT expression, particularly cancers such as breast cancer,
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
XX cancer, pancreatic cancer, cervical cancer, cancers of the central
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be
XX used as hybridisation probes, in chromosome and gene mapping, in
XX chromosome identification and in gene therapy. The present sequence
XX represents a TAT nucleic acid of the invention
XX
SQ Sequence 3127 BP; 629 A; 1001 C; 966 G; 531 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.2e-07 Length: 3127
Score: 302.00 Matches: 157
Percent Similarity: 33.67% Conserved: 78
Best Local Similarity: 22.49% Mismatches: 232
Query Match: 6.29% Indels: 231
DB: 13 Gaps: 27

US-09-920-705-3 (1-901) x ACN37956 (1-3127)
QY 11 AspaSerGlyPheGlnSerAsnLeuTyrValGlySerLeuThrProGluThrThr 30
DB 1112 GAGATGACCAAGCGGCGCACCGCAACCTCTTCAATGTGTAACCTGACCAAGCGTATCT 1171
QY 31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrVal--- 49
DB 1172 GAGGTGAGAGCTCCAAAGGCGCTTGCAGAAATATGACATCATCGAGGAGGTATATCAAG 1231
QY 50 TyrSerSerArgGly-----PheAlaPheIleTyrTyrArgHisValGluGlu 65
DB 1232 AGGCTGCGCCGTGCGCAGGCGGTGCTTATGCTTCAAGTTCAGAACTGGAATG 1291
QY 66 AlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLysIle 85
DB 1292 GCCCATAGGCTTAAGTGGCATGTGCGGCGGAGTGTGGTGCACCCCATTAAGATA 1351
QY 86 GluTyrAlaArgProAlaLysProCysLysSerLeuThrValGlyIleGlyProAsn 105
DB 1352 GGCATGCGAAG---GCCAACCCCAACACATCTGCTTGGGTGGGTGGCTGGAGACTTAC 1408
QY 106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 135
DB 1409 ACGTCACTGGCGGCTGTGCGCCGAGAGTTGACCGCTTGGAGACATTCGACCATTGAT 1468
QY 126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAlaLeuGln 145
DB 1469 CACGTCAAGAGATGACTTGTCTTATATTCATGATCAGAGAGCTTGGACGACGCCAGGCC 1528
```

```

Qy 146 Ala--LysSerMetAenGlyLysProMetGlyGly-----SerPheLeuArgValAsp 162
Db 1529 GCGTGTGCTAAATGAGGGGTTTTCCTTGGGTGAGACAGCCGAGGCTCCGGTGGAT 1588
Qy 163 PheLeuArgSerGlnAlaProLysGluGlnTrpAlaGlySerTrpAsnArgAsn 182
Db 1589 TTGGCCAAAGCAGAG-----1603
Qy 183 GlyAsnMetAsnHisLysProGlnTrpProHisSerTrpGluAspPheLysGlyAspVal 202
Db 1604 -----GAGACTCGGTACCCCGACAGTAC-----1627
Qy 203 GlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGlu 222
Db 1628 CAGCCCTCCGCACTC-----CCTGTGATTATGAGCTGCTCAGATGCA 1672
Qy 223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGlnIleGluArgValLysSerTrpPro 242
Db 1673 TACACCCGCGACCGCAAC-----CTGACGCGGACCTGGGGGAGCAGAGCCGCCA 1726
Qy 243 -----SerArgAsnPheAlaLeuValGluPheArgSerAlaGlu 255
Db 1727 CACCTTCTGTACTAGACCGAGACCGGACTTTTGGAGGGGAGCTGGACCCCGCAGT 1786
Qy 256 GluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLys 275
Db 1787 AAAAGCTCTGACCGCGCAACAGCCTTGAGGCG-----1819
Qy 276 IleMetLysSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTrpSerGly 295
Db 1820 -----TACAGTCCG 1828
Qy 296 MetLysArgSerArgThrAsp-MetPheAsnAsnAspProSerCysValSerSerProHis 315
Db 1829 TCAGTGTGCGACCGCGAGTGTGAGCCTTGGGGGCGAGATGAGAGACCTGGTGGCCAG 1888
Qy 315 AserThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTrpAs 335
Db 1889 C-----CCTGGG-----AAGGAGGCGGAAA 1909
Qy 335 nglyAlaGlyTrpAsnAspValAlaGlyLysGluPro-----AsnTrpArg 350
Db 1910 CGGAGAAAGCCTTTCAGTACCGTGGAGAGCAACCATTCACCATATGAGAAAGGAGT 1969
Qy 350 GArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro-----366
Db 1970 AGGACCAAGGGCAGTGGGCGAGCTCAGACGGGGCTCCGACCGCACCTTGAGCGCAGC 2029
Qy 367 -GlyIleLeuProSerProAlaGlnGlyThrArgArg-----378
Db 2030 CGCAAGAGAAACCACTCCAGTGAAGGACCAAGAGATCCAGACCACTCCCTCAGCAAC 2089
Qy 378 -----378
Db 2090 AGCAGACATGGGGGTGAGAAAGGGGCCACACACACACAGAGGCTGAGACTCT 2149
Qy 379 -----ProMetArgSerAsnPr 384
Db 2150 TCCCAACGGAAGAAAGCAAGACAGACGCGACCATCACCGGACCAAGAGCGGAGCC 2209
Qy 384 GAspSerTrpGluGlyTrpAspProAla-GlnLeuValArgGluSerLysArgThrArg 404
Db 2210 AAGCCTCTGAAAGCCAAACACGAGACCAAAAGCTGAAGATCTTTCAGAGTACGT 2269
Qy 404 rGAspGlySerValAspGlyPheThrProMetGlyValAspGluArg-----419
Db 2270 CAGACACTACAGCTGGGTTGGATGGGCTTGTGTGTAAGAAACAGCTGCTCCACG 2329
Qy 420 -----SerPheGlyArgGlySerValAlaAla-----428
Db 2330 TCTATGATATCTCAGAGGGGAGCAAGGGGTGATCAGACGTCTCAAGACCACT 2389
Qy 429 -----ArgProIleArgGlyPro-ProAsp 436

```

```

Db 2390 TCTGGAGAGAGCTGACCCAGCTGAAGATCCGACGCGCTTCGATGAGACAGCCCAAG 2449
Qy 437 SerAspHisIleTrpArgGlyMetIleAlaLysGlyThrPro-----451
Db 2450 CTTGACGAGATGACACAGCAGCATC-----AAGCAGGGAGGCGCCCAACGGCTATGCGGTC 2503
Qy 452 ValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471
Db 2504 CTTTATGCCAACCGAGCAACCCCGAGTGGCTTGGAGCTGAG--GGATGCCACAGTA 2560
Qy 472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaHisIleTrpAlaValAlaIle 491
Db 2561 GAGCCCGGTCTGACAGAGGCGGCTTCAGAGAACCTC-----2596
Qy 492 GlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTrpGlu 511
Db 2596 -----2596
Qy 512 PheLeuArgTrpLeuSerSerLysAspArgAlaGlyValAlaLysLeuAsp-----528
Db 2597 --GTCTCTACTTGAACAGAGAGCGCGAGGGGTGATCAGCTTGCAGTGGGGGG 2653
Qy 529 -----AspGlyThr--ThrLeuPheLeuValProProSerAspPheLeuThr 543
Db 2654 TCCAGAGGACAGAGCGGACAGGACATGCTTACGCTTCCACCTCCGACCTTTTCCAG 2713
Qy 544 AspValLeuGluValThr-----ArgGlnGluArgLeuTrpGly 556
Db 2714 CAGTACCTTCACAGACACTAAGACATTTGGCGAAGCTAGAGACATGAGATGAGATA 2773
Qy 557 ValValLeuLys-----LeuProProProAlaValProValThrLysSerTrpArg 573
Db 2774 GTCATGTGACAGACATGCTTACCGCAAGCCTGTCTTCCAGCCTCAATGTTGTGTA 2833
Qy 574 GlnGluSer-----GlnSerAsnProLeuHisIleTrp 583
Db 2834 CAAGACATTAATTTAAATCTGATCCCTCTCTACCTTACCATCTT 2881

RESULT 10
AAD53223_4
Continuation (5 of 5) of AAD53223 from base 40001 (Human chromosome 3 p-arm breakpoint
WP Sequence split into 5 fragments LOCUS AAD53223 Accession AAD53223
WP Fragment Name Begin End
WP AAD53223_0 1 110000
WP AAD53223_1 100001 210000
WP AAD53223_2 200001 310000
WP AAD53223_3 300001 410000
WP AAD53223_4 400001 487980

Alignment Scores:
Pred. No.: 2,26e-05 Length: 87980
Score: 302.00 Matches: 157
Percent Similarity: 33.67% Conservative: 78
Best Local Similarity: 22.49% Mismatches: 232
Query Match: 6.29% Indels: 231
DB: 8 Gaps: 27

US-09-920-705-3 (1-901) x AAD53223_4 (1-87980)
Qy 11 AspAspSerGlyPheGlnSerAsnLeuTrpValGlySerLeuThrProGluThrTrp 30
Db 73903 GAGGATGACACAGCGGCGCAGCGCAACCTTCATTTGTAACCTGAGACCAAGCCTATCT 73962
Qy 31 GluSerAspLeuThrGluLeuPheGlyArgTrpGlyAspIleAspArgIleThrVal---49
Db 73963 GAGCTGAGACTCGAAGGCGCTTCAGAAATATGATCATCAGAGAGGTGCTCATCAG 74022
Qy 50 TyrSerSerArgGly-----PheAlaPheIleTrpTrpArgHisValGluGlu 65
Db 74023 AGGCTGCGCGTGGCCAGAGCGGTGCTATGCTTCTCAAGTTCCAGAACTGAGACATG 74082
Qy 66 AlaValAlaLysLeuLysLeuGlnGlyAlaLeuAsnLeuAsnGlySerGlnIleLysIle 85

```

Db	74083	GGCCATAGGGCTTAAGGTGGCCATGTCGGGCGGAGTGAATTGGTCGCAACCCCATTTAAGATA	74147
QY	86	GLUTYRALAARGPROALALAYEPROCYSLYSSERLEUTRVALGILYILGLYPROASN	105
Db	74143	GGCTATGGCAAG---GCCAACCCCCACTGCTCTCGGGGGGTGCCTGGAGCCTAAC	74199
QY	106	VALSERLYASAPRPLEUGLUGLUNHSESERLYSPHEGLYSLILEGLUAPRHEADY	125
Db	74200	ACGTCACTGGCGGCTCTGGCCCGAGAGTTTACCGCTTGGAGACATTCGACCATTTGAT	74255
QY	126	PHLEUAARGLUARGLYETHRALPHEHLEASPTRYTYRGLUMETASPAEALALEUAIN	145
Db	74260	CACGTCAAGAAGATAGCTTTGGCTATATTCATGATACGAGAGCTTGACCGAGCCAGGGC	74319
QY	146	ALA---LYSSERMETASNGILYLYEPROMETGLYLY-----SERPHLEUAARGVALAP	162
Db	74320	GGCTGTGCTAAATATAGGGGTTTTCCTCGGTGAGACAGACCGACGCTCCCGTGAGAT	74379
QY	163	PHLEUAARGSERGINALAPROLYELSGULINTRPALGLYSERTYRASPENARGAAN	182
Db	74380	TTTGGCAAGAAGAGAG---	74394
QY	183	GLYASMETASNHSLYSPROGLINTYRPROHISERTYGLUASPHELYSLYAPVAL	202
Db	74395	-----GAGACTCGGTACCCCGACAGATAC-----	74418
QY	203	GLNPROSERLYVALLEUTRPILEGLYPHEPROTHRALATHRGINCYASNAAPGUL	222
Db	74419	CAGCCCTCGCCACATC-----CCTGTGATTATGAGCTGCTCACAGATGGA	74463
QY	223	GLINILEUHIISASNAIAMELILEUPHEDGLYULILEGLUARGVALLYSERTYRPRO	242
Db	74464	TACACCCCGGACCGCAAC-----CTGAGCGCGCACTCGGTGGGAGACAGACGCCCA	74517
QY	243	-----SERARGASNPHEALALEUVALJLUNHARGSERALAGLU	255
Db	74518	CACCTTCTGTACTAGACCGAGACCGGACCTTTTGGAGGGAGCTGACACAGCCCACT	74577
QY	256	GLUALAARGINCYSLYSGULIYLEUGLNGIYARGNEUPHEASNAAPROHRIILEYS	275
Db	74578	AAAGACTCTGACCCCGAACAAGCCTTGAGGGC-----	74610
QY	276	ILEMETYRSERASNAAPGULNEUPROPROGLINAPASPHISERTHERYTSERGLY	295
Db	74611	-----TACAGTCGC	74619
QY	296	METLYARGSERARGTHINASP-METPHEASNAAPPROSERCYVALISERSEPROHI	315
Db	74620	TCAGTGGCGACCGGAGTGGTGAACGTTGGGGGGCAGATGAGAACCTGTGTTGCCCAAG	74679
QY	315	SEERTHRLIYLEPROGLYSERMETARGPROLEUARGIYTHRASNGLUARGSEPTYRAS	335
Db	74680	C-----CCTGGG-----AAGAGAGCGGGA	74700
QY	335	NGIYALAGLUYRZASNAAPVALVALGILYLSGULPRO-----ASNTTPAR	350
Db	74701	CGGAGAACCTTTCCAGTGAACCGTGGGAGAGACAAACCATTTACCATATAGAGAACGAGT	74760
QY	350	GAARGPROSERALASNGIYTHRGILYLEUAPROSERPROTHNGIYPRO-----	366
Db	74761	AGGACCAAGGCGATGGGACAGTCAAGCGGGGCTCGACCGCACCCCTGAGCGCAGC	74820
QY	367	GLYILEUAPROSERPROLAGINGIYTHRASGAG-----	378
Db	74821	CGCAAGAGAACCACTCCAGTGAAGGAGCCAAAGGAGTCCAGACGACATCTCCTCAGCAAC	74880
QY	378	-----	378
Db	74881	AGCAGACATGGGCTGAGAACGGGGCAACCAACCAACCAAGAGCTGACACTCT	74940
QY	379	-----PROMETARGSERASNP	384

Db	74941	TCCCACGGGAAAGAACAGACAGACGACCGCAATCACCGGACCAACAGAGCCGAGCC	75000
Qy	384	QASperITrpIuglYTYrAepProAla-GInleuValArgGInserTysArgThrArgA	404
Db	75001	AAGCCTCGGAAGAGCCAAACACAGAACCAAAAGCTGAAGATCTTCCAGATACGCT	75066
Qy	404	rgAepGlySerValAepGlyPheThrProMetGlyValAepGluArg-----	419
Db	75061	CAGACACTACAGCTGGGCTGGATAGGCTTCTGAGTGTGAAAACAGCTGCTCCCAAG	75120
Qy	420	-----SerPheGlyArgGlySerValAlaIaIa-----	428
Db	75121	TCTATGATATCTTACAGAGGGGACACAGGGGTATACAGACTCTCTCAAACACACT	75180
Qy	429	-----ArgProIleArgGlyPro-ProAep	436
Db	75181	TCTGGAGCAAGCTGACCACTGAGAGTGCACGACGCTTGCATCTGACACAGCCCAAG	75240
Qy	437	SerAepHisIleTPrArgGlyMetIleAlaValGlyGlyThrPro-----	451
Db	75241	CTTACGAGGTGACACACAGCGATC-----AACAGAGGAGGCCCAACGGCTATGGGTC	75294
Qy	452	ValCysCysValArgCysValProMetGlyTysGlyIleGluThrTyrSerProGluVal	471
Db	75295	CTCTTACGACCCACAGCAACCCCAAGGCGCTTGACCATGAG-----CGATGCCACAGTA	75351
Qy	472	ValAsnCySerAlaArgThrAepIleAsnMetLeuAlaValHisTyrAlaValAlaIle	491
Db	75352	GAGCCCGGTCTGCAGAGCGCGCTTCTCAGAACCTG-----	75387
Qy	492	GlyCysGluIleValPhePheValProAepArgGluGluAepPheAlaSerTyrThrGlu	511
Db	75387	-----	75387
Qy	512	PheIleuArgTyrIleuSerSerTysAepAlaGlyAlaGlyValAlaValLeuAep-----	528
Db	75388	--GTCCTCTACTTGGAAACAGAAAGCAGCGCGAGGGGTATCAGCTTGGACATGGGGGGG	75444
Qy	529	-----AepGlyThr-----ThrLeuPheValProProSerAepPheLeuThr	543
Db	75445	TCCAAAGGACAGAGACGGGACAGGACATGCTCTACGCTTCCACCTCGGACCTTTCCAG	75504
Qy	544	AspValLeuGlnValThr-----ArgGlnGluArgLeuTyrGly	556
Db	75505	CAGTACTCTCAGCTCAGACCTAAGACATTGGCGAAGCTAGAAAGAAACACATGGTGATA	75564
Qy	557	ValValLeuLys-----LeuProProProAlaValProValThrAlaSerTyrArg	573
Db	75565	GTCATCTGTCAGAGACACTGCTGACCCAAAGCTGCTTTCCACAGGTCATGTTGTGTCA	75624
Qy	574	GlnGlnSer-----GlnSerAsnProLeuHisTyr	583
Db	75625	CAAAAGACATTATTTAATCTGATCCCTCTCTACCCCTACCACTTT	75672
RESULT 11			
ID	ADS95035	standard, DNA, 2673 BP.	
XX	ADS95035;		
DT	02-DEC-2004	(first entry)	
DE	Human therapeutic DNA sequence #74, SEQ ID 74.		
KW	Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;		
KW	Gene therapy; Proliferative disorder; cancer; inflammatory disorder;		
KW	immune disorder; bacterial disorder; viral disorder; human; gene; ds.		
OS	Homo sapiens.		
XX			
FN	WO2004039319-A2.		
XX			
PD	13-MAY-2004.		


```

Db      988 GAGAGTACACGAGGCGGCGACGCGCAACCTTCATTGTAAGTGAACGACGAGTATCT 1047
QY      31 GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrVal 49
Db      1048 GAGGTGAGGTGCGAAGGCGCTTCGAGAAATATGATGCAATCGAGGAGGTGATCAAG 1107
QY      50 TyrSerSerArgGly-----PheIlePheIleTyrTyrArgHisValGluGlu 65
Db      1108 AGGCTGCGCGGTGCGGCGGCGGTGCTGCTTATGCTTCAAGTTCCAGAACTCGAGCATG 1167
QY      66 AlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLeuVal 85
Db      1168 GCCCATAGAGGCTTAAGGTGCGCATGTGCGGCGGAGTATGTTGTCGCAACCCCATTAAGATA 1227
QY      86 GluTyrAlaArgProAlaLysProCysLysSerLeuTyrValGlyGlyIleGlyProAsn 105
Db      1228 GGGTATGGCGAG--GCCAACCCCGACCACTGCTCTGGGTGGTGGCTGGGACCTTAAC 1284
QY      106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 125
Db      1285 ACCTCACTGGCGGCTTGCGCCGAGAGTTGACCGCTTTGGAGCATTCGAGACCATTAAT 1344
QY      126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
Db      1345 CAGCTCAAGAGATAGCTTGGCTTATATTCAGTACGAGAGCTTGACGCGACGCCAGGCC 1404
QY      146 Ala--LysSerMetAsnGlyLysProMetGlyGly-----SerPheLeuArgValAsp 162
Db      1405 GCGTGTCTAAATATGAGGGGTTCCTTGGGTGAGCAAGACCGCAGGCTCCCGTGAT 1464
QY      163 PheLeuArgSerGlnAlaProLysLysGluGlnThrAlaGlySerTyrAspAsnArgAsn 182
Db      1465 TTTCGCAAGCGAG-- 1479
QY      183 GlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspVal 202
Db      1480 -----GAGACTCGGTACCCCGACAGTAC----- 1503
QY      203 GlnProSerLysValLeuTyrIleGlyPheProProThrAlaThrGlnCysAsnAspGlu 222
Db      1504 CAGCCCTCGCACTC-----CTGTGCATTATGAGCTGCTCAGCATGGA 1548
QY      223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrPro 242
Db      1549 TACACCCGCGACCGCAAC-----CTGAGCGCGGCTGTGCGGACAGAGCGCCCA 1602
QY      243 -----SerArgAsnPheAlaLeuValGluPheArgSerAlaGlu 255
Db      1603 CACCTTCTGTACTCAGACCGAGACCGGACTTTTGGAAAGGAGACTGACGACGCCCACT 1662
QY      256 GluAlaArgGlnCysLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLys 275
Db      1663 AAAAGCTCTGACCGCGGAAACAGCTTGAGGGC----- 1695
QY      276 IleMetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGly 295
Db      1696 -----TACAGTCCG 1704
QY      296 MetLysArgSerArgThrAsp-MetPheAsnAsnAspProSerCysValSerSerProHis 315
Db      1705 TCAGTGGCGACCGGAGGTGTGAGCGTTGGGGGCGAGATGAGAACCTGTGTTGCCCAAG 1764
QY      315 sSerThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAs 335
Db      1765 C-----CTGGG-----AAGAGAGCGCGGAAA 1785
QY      335 ngIylalagIuTyrAsnAspValalGlyLysGluPro-----AsnThrArg 350
Db      1786 CGGAGAGGCTTTCACGTGACCGTGGAGGAGCAACCATTCACCATATGAGGAACGAGGT 1845
QY      350 gatPProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro----- 366

```

```

Db      1846 AGGACCAAGGCGAGTGGGAGCAGTCAAGCGGGGCTCCGACCGCACCCCTGAGCGCAG 1905
QY      367 -GlyIleLeuProSerProAlaGlnGlyThrArgArg----- 378
Db      1906 CGGAGAGAACCACTCCAGTAAAGGAGCAAGAGATCGACGACAACTCCTCAGCAAC 1965
QY      378 ----- 378
Db      1966 AGCAGCATGGGCTGAGGAAGGGGCCACACACACACACACAGAGCTGCAACTCT 2025
QY      379 -----PrometArgSerAsp 384
Db      2026 TCCCAAGGAGAGAGCAAGACAGCGCGCAATCACCGGACCAAGAGGCGGAGGCC 2085
QY      384 oAspSerTyrGluGlyTyrAspProAla-GlnLeuValArgLysSerLysArgThrArg 404
Db      2086 AAGCCTCTGGAAGAGCCAAACAGACGACCAAAAGCTGAAGATTTTCAGAGTACGCT 2145
QY      404 rGAspGlySerValAspGlyPheThrProMetGlyValAspGluArg----- 419
Db      2146 CAGACACTACAGCTGGGTTGAAATGGGCTTCGTGTGTGAAAAACAGTGTCTTCCCAAG 2205
QY      420 -----SerPheGlyArgGlySerValAlaAla----- 428
Db      2206 TCTATGATATCTCTAGAGGGGAGCCAGGGGGTATCAGACAGTCTCTCAAGACACACT 2265
QY      429 -----ArgProIleArgGlyPro-ProAsp 436
Db      2266 TCTGGAGACAGCTGACCCAGCTGAAGATCGCCAGCGCTTCGACTGACAGCAGCCAG 2325
QY      437 SerAspHisIleThrArgGlyMetIleAlaLysGlyGlyThrPro----- 451
Db      2326 CTTCAGAGGTCAACAGACGATC-----AGCAGGGAGGCCCAACGCGCTATGCGGCTC 2379
QY      452 ValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471
Db      2380 CTTTATAGCACCGACGCAACCCCGATGGGCTTGACAGTGAAG--CGGATGCCACAGTA 2436
QY      472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIle 491
Db      2437 GAGCCCGTCTCAGAGCGCGGCTCTCAGAACTG----- 2472
QY      492 GlyCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGlu 511
Db      2472 ----- 2472
QY      512 PheLeuArgTyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAsp----- 528
Db      2473 ---GTCTCTACTTGAACAGAAAGCAGCGCGGAGGGGTGATGAGCTTGCAGTGGGGGG 2529
QY      529 -----AspGlyThr--ThrLeuPheLeuValProProSerAspPheLeuThr 543
Db      2530 TCCAAAGGAGAGAGCGGACAGGCAATGCTTACGCTTCCACCCGTGCACTTTTCCAG 2589
QY      544 AspValLeuGln 547
Db      2590 CAGTACCTCCAG 2601

```

RESULT 13

ABL05395 standard; cDNA; 3084 BP.

ABL05395;

26-MAR-2002 (first entry)

Drosophila melanogaster expressed polymucleotide SHQ ID NO 10667.

Drosophila; developmental biology; cell signalling; insecticide;

pharmaceutical; gene; ss.

Drosophila melanogaster.

PN WO200171042-A2.
 XX
 XX 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR P-PSDB; ABB61292.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 XX
 PS Claim 1; SEQ ID NO 10667; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
 CC AB57072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SX Sequence 3084 BP; 706 A; 868 C; 885 G; 625 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3,6e-06	Length:	3084
Score:	287.50	Matches:	140
Percent Similarity:	32.70%	Conservative:	65
Best Local Similarity:	22.33%	Mismatches:	201
Query Match:	5.99%	Indels:	221
DB:	4	Gaps:	27

US-09-920-705-3 (1-901) x ABL05395 (1-3084)

QY	11	AspAspSerGlyPheGlnSerAsnLeuTyrPValGlySerLeuThrProGlnThr	30
DB	1111	GAGGATGATCCTCTTCTACGCGCACACTGTTGCGAGAAATCTGGAGTTTCCATCCG	1170
QY	31	GluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAspArgIleThrValTyr	50
DB	1171	GACGAGCAACTGTCGATCTTTGGCAATACGAGGTGATGATGACATCAGCATCAAG	1230
QY	51	-----SerSerArgGlyPheAlaPheIleTyrTyrArgHisValGluLeu	65
DB	1231	CGACGCCGCCCGGATGCGGGAACGCGTTGCGGTGATGCTTACCAAACTCGATATG	1290
QY	66	AlaValAlaAlaLysGlnLalaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLysIle	85
DB	1291	GCTCACCGAGCGGAAGATCGAATTTGTCGTGACAGTACATTGGCAATTTCAGCGAAGATC	1350
QY	86	GluTyrAlaArgProAlaLysProCysLysSerLeuTyrPValGlyIleGlyProAsn	105
DB	1351	GGATACGGGAAA---GTGACTCCGGCAACTCGATGATGATGATGAGGCTTATGCTCTGG	1407
QY	106	ValSerLysAspAspLeuGlnGluPheSerLysPheGlyLysIleGluAspPheArg	125
DB	1408	ACTTCTGTACTCACTGAGCGAGAGGTTGATCTCTTCGCGCATCAAGAAATTGAA	1467
QY	126	PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln	145
DB	1468	TACCAAGAGGGCGGACCGTATATATTCAGTACAGACAGTGTGAGGACGCCGCGG	1527

QY	146	Ala---LysSerMetAsnGlyLysProMetGlyGly-----SerPheLeuArgValAsp	162
DB	1528	GGGTAAAGAGATGCGAGAGATTCCATTTGGGTGAGCGAGCTCGACTCCGACGAGAC	1587
QY	163	Phe-----LeuArgSerGlnAlaProLysLys	171
DB	1588	TTTGGCGAGTTGCCCGGCGCCACACCGGACGCGCTTTCAAAGAGTCAAAAGCGCCCATAT	1647
QY	172	GluGlnTyrPValGlySerTyrAspAsnArgAsnGlyAsnMetAsnIleLysProGlnTyr	191
DB	1648	GATGAAGGCGCTTGAGATAC-----AGACGTCGGGAATAC	1683
QY	192	ProHisSerTyrGluAspPhe-----LysGlyAspValGlnProSerLys	206
DB	1684	GATCCGATTACGAAGAGTCGCGCTACGCTCCGCGGCTGCTATTCCTCG-----	1737
QY	207	ValLeuTyrPValGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHis	226
DB	1738	-----TATCCACTTAGGGGT-----	1752
QY	227	AsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPhe	246
DB	1753	-----GGCTACCGTGCAGCTGGG---	1770
QY	247	AlaLeuValGluPheArgSerAlaGluGlnAlaArgGlnCysLysGluGlyLeuGlnGly	266
DB	1771	-----GGATATCGCGGA	1782
QY	267	ArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProProGlu	286
DB	1783	AGA-----GGAAAGGCGATGATCATTTATCAACAGACGTCGATAGGCCG---	1827
QY	287	GlnAspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsn	306
DB	1827	-----	1827
QY	307	AspProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArg-----	324
DB	1828	-----CCGATCTCTGCTCTTACGAGGATCTCTTCTTTCAGTT	1866
QY	325	ProLeuArgGlyThrAsnGluArgSerTyrAsnGlyValaGluTyrAsnAspValaGly	344
DB	1867	CCCCCTCCGCGC-----GGA	1881
QY	345	LysGluProAsnTyrPArgArgProSerAlaAsnGly-----ThrGlyIleLeuProSer	362
DB	1882	GTCGAGATGAGTGCACGCTCCGCGCAGGAATCTGACAGACAGAGGCCGCTCCAGC	1941
QY	363	ProThrGlyProGlyIle-----LeuProSerProAlaGlnGlyThrArgArg	378
DB	1942	TCTGGGAACTGCTGTGAAGACGCTCTCGTTCTAGATCCCATTTGAAGCGGCTCGCTCT	2001
QY	379	Pro-----MetArgSerAsn-----	383
DB	2002	CCCGATCTGATTTCTGACACCTCGACTCGTGGAAACAGCCCTTGATCAGCAAGTACG	2061
QY	384	-----ProAsp-----SerThrGluGlyTyrAspProAlaGlnLeu	395
DB	2062	GTTCCAAATGTCGCGCGCAAGTGCAGACAGCTGTGACCGGA-----GGGCTAATC	2112
QY	396	ValArgLysSer-----LysArgThrArgAspArgLysSerValaLysGlyPhe	411
DB	2113	CTCAAGAGCTCACTGTTCCAGCAAGTTTCAATTAGACGAGACGAGATACGATATTGTG	2172
QY	412	ThrProMetGlyValaAspGluArgSerPheGlyArgGlySerValaAlaAlaArgProIle	431
DB	2173	GAGTCGCTGATGCGCGATGAGGAGGCGAAGCAACACTAGTATCACCCAGCAGATGCGGT	2232
QY	432	ArgGlyProProAspSerAspHisIleTyrArgGlyMetIleAlaLysGlyThrPro	451
DB	2233	CTGATCTCGCCCAAGCTGCGACGTCGCAAAAGTATTCGACTTCATCG-----	2283
QY	452	ValCysAlaArgCysValaProMetGlyLysGlyIleGluThrLysLeuProGluVal	471

```

Db      2284 -----:::|||||
          TCGACGCGCATTTTCATGCTCTGCGAGGCTCCACA-----AACGAC 2325
QY      472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaIshIstYrAlaValAlaIle 491
          |||||
Db      2326 ACCAATTGCG----- 2334
QY      492 GlyCyGluIleValPhePheValProAspArgGluGlnAspPheAlaSerTYrThrGlu 511
          |||||
Db      2335 -----GACGATGCCAGCGCTGACAGCGCA 2358
QY      512 PheLeuArg-----TYrLeuSerSerIleAspArgAlaGlyValAlaIleValLeu 527
          |||||
Db      2359 CCACCTAAGAAACTTGGTTCTCTTACGACAGAGAGCGCGCTGCGTGAATCTGCTG 2418
QY      528 AspAspGlyThrThr-----LeuPheLeuValProProSerAspPheLeu 542
          |||||
Db      2419 CTAAACAGAGAGACGAGCACTGGGCTGCTGACGCTTCCCTCCCTCGACTTTTTC 2478
QY      543 ThrAspValLeuGlnValThr-----ArgGlnGluArg 553
          |||||
Db      2479 ACAGAGCTGCTAAGCGCAGCTTGTACAGTGTGACGAGAGAGGCGCTGAAGAGATCAC 2538
QY      554 LeuTYrGlyValValLeuIys 560
          |||||
Db      2539 CTGCTGATTTGTGTGTGCGCC 2559

RESULT 14
ABL05394/c
ID      ABL05394 standard; cDNA; 9744 BP.
XX      ABL05394;
XX      26-MAR-2002 (first entry)
XX      Drosophila melanogaster expressed polynucleotide SEQ ID NO 10664.
XX      Drosophila; developmental biology; cell signalling; insecticide;
XX      pharmaceutical; gene; ss.
XX      Drosophila melanogaster.
XX      WO200171042-A2.
XX      27-SEP-2001.
XX      23-MAR-2001; 2001WO-US009231.
XX      23-MAR-2000; 2000US-0191637P.
XX      11-JUL-2000; 2000US-00614150.
XX      (PEKE ) PE CORP NY.
XX      Venter JC, Adams M, Li PWD, Myers EW;
XX      WPI, 2001-656860/75.
XX      P-PsDB; ABB61291.
XX      New isolated nucleic acid detection reagent for detecting 1000 or more
XX      PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX      PT interactions.
XX      Claim 1; SEQ ID NO 10664; 21p + Sequence Listing; English.
XX      The invention relates to an isolated nucleic acid detection reagent
XX      CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX      CC useful in developmental biology and in elucidating cell signalling and
XX      CC cell-cell interactions in higher eukaryotes for the development of
XX      CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX      CC discloses genomic DNA sequences (AB16176-AB10511), expressed DNA
XX      CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
XX      CC AB572072). The sequence data for this patent did not form part of the
XX      CC printed specification, but was obtained in electronic format directly

```

```

CC      from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX      SQ      Sequence 9744 BP; 2598 A; 2274 C; 2331 G; 2541 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.:      1, 24e-05      Length:      9744
Score:          287.50      Matches:      140
Percent Similarity:      32.70%      Conservative:      65
Best Local Similarity:      22.33%      Mismatches:      201
Query Match:      5.99%      Indels:      221
DB:              4      Gaps:      27

US-09-920-705-3 (1-901) x ABL05394 (1-9744)
QY      11 AspAspSerGlyPheGlnSerAsnLeuTyrValGlySerLeuThrProGluThrThr 30
          |||||
Db      7411 GAGATGATGCTCTTCTTACGCGCACACTGTTTGACGAAATCTGAGGTTACATCGCC 7352
QY      31 GluSerAspLeuThrGluPheGluArgTYrGlyAspIleAspArgIleThrValTYr 50
          |||||
Db      7351 GACGACGAATGCTGCTGCACTTTTGCAATACGAGTGTGTGATGATGACATCGACATCAAG 7292
QY      51 -----SerSerArgGlyPheAlaPheIleTYrTYrArgHisValGluGln 65
          |||||
Db      7291 CGACCGCGCGGCTACGGGGAACGGCTTCGGGTTTGATCGTTACCAAAACCTGATATG 7232
QY      66 AlaValAlaIleValGluIleValLeuGlnGlyAlaLeuLeuAsnGlySerGlnIleValIle 85
          |||||
Db      7231 GCTCACCGAGCGAAGATCGAATGTCGTGACAGTACATTCGCAAGTTTCAGTGCAGAGATC 7172
QY      86 GluTYrAlaArgProAlaIleProCysIysSerLeuTyrValGlyIleGlyProAla 105
          |||||
Db      7171 GGATACGGGAAA---GTGACTCCGCACTCGTATGTGATCGAGCGCTTAGGTGCTGG 7115
QY      106 ValSerIleAspAspLeuGluGluGluPheSerIlePheGlyIleGluAspPheArg 125
          |||||
Db      7114 ACTTCTGTCACTCAGCTGAGGAGAGGTTGATCGCTTTCGTCGATCAAGAAATGGA 7055
QY      126 PheLeuArgGluArgIleThrAlaPheIleAspTYrTYrGluMetAspAspAlaLeuGln 145
          |||||
Db      7054 TACCAAGAGGCGAGCGCTGATTCCTTATTCAGTACGAGACAGTGCAGCGACGCGCG 6995
QY      146 Ala---LysSerMetLeuAsnGlyIysProMetGlyGly-----SerPheLeuArgVala 162
          |||||
Db      6994 GCGTAAAGAGATGCGAGGATTCCTATTGGGTGACCGAGCGTCGATCGCACGAGC 6935
QY      163 Phe-----LeuArgSerGlnAlaProIleIys 171
          |||||
Db      6934 TTTCGCGAATGCCCGCGCCACACCGGACGCCCTTTCAGACAGCTCAAGCGCCATAT 6875
QY      172 GluGlnTyrAlaGlySerTYrAspAsnArgAsnGlyAsnMetAsnHisIleYsProGlnTYr 191
          |||||
Db      6874 GATGAAGTGCCTTGAGTAC-----AGAGCTCCGGAATAC 6839
QY      192 ProHisSerTYrGluAspPhe-----LysGlyAspValGlnProSerIys 206
          |||||
Db      6838 GATCCGTAATACGAAGATCCCGCCCTACGCTCCGCGGTGCTATTCTCCG----- 6785
QY      207 ValLeuTyrIleGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHis 226
          |||||
Db      6784 -----TATCCACTAGGGGT----- 6770
QY      227 AsnAlaMetIleLeuPheGluIleGluArgValIysSerTYrProSerArgAsnPhe 246
          |||||
Db      6769 -----GGCTACCGTGAAGTGGG----- 6752
QY      247 AlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysIleGluIleGluGlnGly 266
          |||||
Db      6751 -----GGAATATCGCGGA 6740
QY      267 ArgLeuPheAsnAsnProArgIleIleYsIleMetTYrSerAsnAspGluLeuProProGln 286
          |||||
Db      6739 AGA-----GGAAGAGCATGTACATTATCACAGAGCATGTAGCGCG----- 6695

```

```

QY 287 GlnAspAspThrSerPheTyrSerGlyMetIleYargSerArgThrAspMetPheAsnAsn 306
Db 6695 -----
QY 307 AspProSerCysValSerSerProHisSerThrGlyIleProGlySerMetArg----- 324
Db 6694 -----
QY 325 ProLeuArgGlyThrAsnGluArgSerTyrAsnGlyValAsnIleValAsnValGly 344
Db 6655 CCCCCTCCGGC-----
QY 345 LysGluProAsnTrpArgArgProSerAlaAsnGly-----ThrGlyIleLeuProSer 362
Db 6640 GTGAGAGATGATGGGCGACGCCCGCAGAGATCGTACAGACAGAGAGCCGCTCCAGC 6581
QY 363 ProThrGlyProGlyIle-----LeuProSerProAlaGlnGlyThrArgArg 378
Db 6580 TCTGGGAACTGTGTAGAAACGCTCTCTTCTAGATCCCATTTGAAGCGGCTCGCTCT 6521
QY 379 Pro-----MetArgSerAsn----- 383
Db 6520 CCGGATCTGATTTCTGACACTCGTCTGAGACGACGCCCTTGCAATCAGCAGTACAG 6461
QY 384 ---ProAsp-----SerTrpGluGlyTyrAspProAlaGlnLeu 395
Db 6460 GTTCCAGATGTGGCGCGCAAGTGCAGACAGTGTGACCGGA-----GGCTAATAC 6410
QY 396 ValArgGluSer-----LysArgThrArgArgAspGlySerValAspGlyPhe 411
Db 6409 CTCAGAGCTCACTGTTCCAGCAGCAATTTCATTTGACGACGAGATACGAGATTTGG 6350
QY 412 ThrProMetGlyValAspGluArgSerPheGlyArgGlySerAlaIleAlaArgProIle 431
Db 6349 GAGTCCCTATATGGCCATGAGAGGCGCAACACCTTCATTCACCCAGACGACTGGCT 6290
QY 432 ArgGlyProProAspSerAspHisIleTrpArgGlyMetIleAlaArgGlyGlyThrPro 451
Db 6289 CTGGATCCCGCCCAAGCTGACGACGTGCAAAAACGTAATTCATCCATCG----- 6239
QY 452 ValCysCysAlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluVal 471
Db 6238 -----TCGACGCCCATTTTCATGGGTCTGGCAGGGTCCACA-----AACGAC 6197
QY 472 ValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaLysIleTyrAlaValAlaIle 491
Db 6196 ACCAATTGC----- 6188
QY 492 GlyCysGluIleValPhePheValProAspArgGluAspPheAlaSerTyrThrGlu 511
Db 6187 -----GACGATGCCAGCGCTGCAGACGGGA 6164
QY 512 PheLeuArg-----TyrLeuSerSerLysAspArgAlaGlyValAlaLysLeu 527
Db 6163 CCACTAAGAACTTGTTCTTCTATCTTAAGCAGAAAGAGCGCGCTGCTGATCTCGCTG 6104
QY 528 AspAspGlyThrThr-----LeuPheLeuValProProSerAspPheLeu 542
Db 6103 CTTAACAAGAGACGAGAGCAACTGGGGTCTGTACGCTTCCCTCCCTGCAGCTTTTCC 6044
QY 543 ThrAspValLeuGlnValThr-----ArgGlnGluArg 553
Db 6043 AAGAGAGCTGCTAAGAGCACTTGTCACAGTCTGACGAGAGAGCGCTGAAGAGAGATCAC 5984
QY 554 LeuTyrGlyValValLeuLys 560
Db 5983 CTGGTATTTGTGTGTGTGCGC 5963

```

RESULT 15
 ABZ21697
 ID ABZ21697 standard; cDNA; 3383 BP.
 XX

```

AC ABZ21697;
XX
DT 27-FEB-2003 (first entry)
XX
DE Human RBM15-L protein encoding cDNA SEQ ID NO:11.
XX
KW Human; RBM15; RNA binding motif protein 15; megakaryoblastic leukemia 1;
KM MKL1; fusion protein; acute megakaryoblastic leukemia; AMKL; cytostatic;
KW t(1; 22) chromosomal rearrangement; gene therapy; chromosome 1p13; gene;
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 84..2957
FT /tag=a
FT /product="RBM15-L protein"
XX
PD WO200288309-A2.
XX
PD 07-NOV-2002.
XX
PF 23-APR-2002; 2002WO-US012797.
XX
PR 27-APR-2001; 2001US-0286910P.
XX
PA (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
XX
PI Morris SW, Ma Z, Hitzler JK;
XX
DR WPI; 2003-103455/09.
DR P-PSDB; ABB56070.
XX
PT New RNA-binding motif protein-15 (RBM15)-megakaryoblastic leukemia-1
PT (MKL1), MKL1-RBM15-S and MKL1-RBM15-S+AR fusion proteins, useful for
PT identifying agents useful for treating patients with acute
PT megakaryoblastic leukemia.
XX
PS Disclosure; Page 89-93; 109pp; English.
XX
CC The present invention describes an RNA-binding motif protein-15 (RBM15)-
CC megakaryoblastic leukemia-1 (MKL1) fusion protein, a MKL1-RBM15-S fusion
CC protein, and a MKL1-RBM15-S+AR fusion protein associated with acute
CC megakaryoblastic leukemia (AMKL). Also described (1) an antibody that
CC specifically binds to the RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AR
CC fusion proteins; (2) a non-human transgenic animal that has been altered
CC to express a gene encoding a RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AR
CC fusion protein; (3) identifying an agent capable of binding to a RBM15-
CC MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AR fusion protein; (4) detecting the
CC t(1;22) chromosomal rearrangement associated with AMKL; and (5) screening
CC for agents capable of (selectively) inhibiting the activity of a fusion
CC protein arising from the t(1;22) chromosomal rearrangement associated
CC with AMKL. The fusion proteins have cytostatic activity and can be used
CC in gene therapy. The RBM15-MKL1, MKL1-RBM15-S, or MKL1-RBM15-S+AR fusion
CC proteins and nucleotide molecules are useful for designing and preparing
CC agents that specifically inhibit the expression of the RBM15-MKL1 or MKL1-
CC RBM15 genes in cells for therapeutic and other purposes. The transgenic
CC animals are useful for identifying and testing carcinogenic or
CC therapeutic compositions. The methods are also useful for detecting the
CC t(1;22) chromosomal rearrangement associated with AMKL, or for
CC identifying agents useful for treating patients with AMKL. The antibodies
CC can be used to selectively kill cells expressing RBM15-MKL1, MKL1-RBM15-
CC S, or MKL1-RBM15-S+AR fusion proteins. RBM15 is located to chromosome
CC 1p13, and MKL1 is located to chromosome 22q13. The present sequence
CC encodes the human RBM15-L protein which is given in the exemplification
CC of the present invention
XX
SQ Sequence 3383 BP; 812 A; 854 C; 952 G; 765 T; 0 U; 0 Other;

```

Alignment Scores: 1.53e-05 Length: 3383
 Pred. No.: 276.50 Matches: 162
 Score:

Percent Similarity: 35.87%
 Best Local Similarity: 23.24%
 Query Match: 5.76%
 DB: 8
 Gaps: 34

US-09-920-705-3 (1-901) x ABZ21697 (1-3383)

QY ProPhaArg-----AlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
 DB CTTTTCAGAGAGATGATGATGATTTTCAACCCAGAGATGATCAGAGCTAACCGAGCGCTC 1211
 QY 1152 CTTTTCAGAGAGATGATGATGATTTTCAACCCAGAGATGATCAGAGCTAACCGAGCGCTC 1211
 DB 21 TrpValGlySerLeuThrProGluThrGlnSerAspLeuThrGlnLeuPheGlyArg 40
 DB 1212 TTTCTGGGCAACTAGACATCACTGTAACGAGAGATGATTTAAGAGGGGCTTGAATCGC 1271
 QY 41 TyrGlyAspIleAspArgIleThrVal--TyrSerSerArgGly-----Phe 55
 DB 1272 TTTGGAGTCATCAAGAGATGATATCAGAGCGCTTCTCGCGCCAGACTAGACTATAC 1331
 QY 56 AlaPheIleTyrThrArgHisValGlnGluAlaValAlaValSerGlnAlaLeuGlnGly 75
 DB 1332 GGCCTTCTCAATTTGAGAACTTAGATATGCTCACCGGCAAAATTAGCAATGCTCTGC 1391
 QY 76 AlaAsnLeuAsnGlySerGlnIleLeuTyrAlaArgProAlaLysProCysLys 95
 DB 1392 AAAATTATATATGGGAATCTATCAAAATGGTTATGCTAA--GCTACACCCACCACC 1448
 QY 96 SerLeuThrValGlyGlyIleGlyProAsnValSerLysAspAspLeuGlnGluPhe 115
 DB 1449 CGCCTCTGGGTGGAGGCGCTTGAGCCTTGCGTCTCTGCTCGCCGACAGAAATTT 1508
 QY 116 SerLysPheGlyLysIleGlnAspPheArgPheLeuArgGlnArgLysThrAlaPheIle 135
 DB 1509 GATGATTTGGCAACCAATGACCACTACGAAAGGTATGTTGGCATATATC 1568
 QY 136 AspTyrTyrGlnMetAspAspAlaLeuGlnAla--LysSerMetAsnGlyLysProMet 154
 DB 1569 CAGTATGAAGAAGCCTGGATGCGAGCGCATGCTGCTGAGCCCATATGCGGGCTTCCCACTT 1628
 QY 155 GlyGlyLys-----SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysGln 172
 DB 1629 GGTGGCCCAAGATCAGAGCCTTAGATGATGACTTTGGCCGACACCAATCTCTTACAGAG 1688
 QY 173 GlnTrp-----AlaGlySerTyrAspAsn 180
 DB 1689 CAGTATCTGCAGCCTTGTGCTGACTATGATGAGCTGTGACAGATGCTTTTGGACAT 1748
 QY 181 Arg-----AsnGlyAsnMetAsnHisLysProGlnIleTyrProHisSerTyr 195
 DB 1749 CGGGACACAGACCCCTTGAAGGGGTGCTCGGATAGAGACACA-----CCCTTACTATAC 1802
 QY 196 GlnAspPheLysGlyAspValGlnProSerLysValLeuTrpIleGlyPheProThr 215
 DB 1803 AGAGATCGATAGAGACCTTATCTGACTCGAT--TGGGTGCCACCCCAACCCCA 1859
 QY 216 AlaThrGlnCysAsnAspArgGlnIleLeuHisAsnAlaMetIleLeuPheGlnGluIle 235
 DB 1860 GTCCGA-----GAACGACAGACTCGGACTGCACTTCTGTGCTGCTTAT 1907
 QY 236 GlnArgValLysSerTyrProSerArg-----AsnPheAlaLeuValGlnPheArgSer 253
 DB 1908 GAGCCACTGATATGCTCATGATTCGACGGCGGATGCTTGTGCTTGGACCGGACAGAGGT 1967
 QY 254 AlaGlnGluAlaArgGlnCysLysGlnGlyLeuGlnGlyArgLeuPheAsnProArg 273
 DB 1968 GATCGAGATCTGCCAGACAGAGAC-----CAGCCTAGG 2003
 QY 274 IleLysIleMetTyrSerAsnAspGlnLeuProGlnIleAsp-----Asp 289
 DB 2004 AACGGA-----AGCTGCTGAGAGAGTGCAGAGACTCATCTCGAT 2045
 QY 290 ThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAspProSer 309

DB 2046 AGGTCTCTGAGAGTACCGCCACGAAACGT-----CAC 2081
 QY 310 CysValSerSerPro--HisSerThrGlyIleProGlySerMetArgProLeuArgGly 328
 DB 2082 TGGCTCTCTTCTCCGACCGCAGTCCGAATTGAGAGATGAGCGGATGTTACACACG 2141
 QY 329 ThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsnAspValAlaGlyLysGluProAsn 348
 DB 2142 GACATATGATCGATCT-----TCCGCTTCTCTTGGAAAGCCCTCT 2183
 QY 349 TrpArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIle 368
 DB 2184 -----CCATCAGAACGAGACGAGGTAGTTG----- 2210
 QY 369 LeuProSerProAlaGlnIleThrArgArgProMetArgSerAsnProAspSerTrpGlu 388
 DB 2211 -----GAGAAAGCCAGGAGTACAGACGAGAC--CGTAAAACTCTCATCAGCTGAA 2261
 QY 389 GlyTyrAspProAlaGlnLeuValArgGlnSerLysArgThrArgAspArgLysSerVal 408
 DB 2262 -----CGAGATGAGAACACCGGACAACTGCTCCACT 2294
 QY 409 AspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGly-----Ser 425
 DB 2295 GAGGAAAAAGCCCTCTGAAAAAGAACCCGCTGATGGAGTGCACCTAGACACACG 2354
 QY 426 ValAlaAlaArgProIleArgGlyPro----- 434
 DB 2355 ACTGCTCTCCAGAGTGAAGTCCCGTCCAGAAACAGATGGGGGAGACGCCCTGTG 2414
 QY 435 -----ProAspSerAspHisIleTrpArgGlyMet----- 444
 DB 2415 GCATCAGCTCTCCAAACTCTGTTGGCTGCGAGGGATGCTTACTAGAAACACG 2474
 QY 445 -----IleAlaLysGlyGlyThrProValCysCysAlaArgCys 457
 DB 2475 AACTTCTTCCAAACATGATGATGTTGAGAGGAGACCTCCAAATGGCTAGTACTTCTT 2534
 QY 458 Val-----PrometGlyLysGlyIleGlnThrLysLeuProGluValAlaAsnCys 474
 DB 2535 GTGGAGGTTCAACTGGAGGCAAGTGGCCAGCTCAGATGATCACTCAGGCTCCCGTTGG 2594
 QY 475 SerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAla-----IleGly 492
 DB 2595 GAC--CAGCCCAAGTTGATGATGATGATCTCAACGATCAAAAGTACAGAGGCCCAATGAT 2651
 QY 493 CysGluIleValPhePheValProAspArgGlnGlu----- 504
 DB 2652 TATGCCATCTTGTGGCTGTGCTGAGATGTTGACAGCCGGTCCCTCTTCTGACGT 2711
 QY 505 -----AspPheAlaSerTyrThrGlu-----PheLeuArgTyrLeuSer 517
 DB 2712 GATCAGACAGTCCCACTTCTACTCAGAGGCCACTTAGAGAACTGTGCTATTTAAAG 2771
 QY 518 SerLysAspArgAlaGlyValAlaLysLysAspAspArgGlyThrThrLeuPheLeuValPro 537
 DB 2772 CAAGACAGCAAGCCGGGTGATGACGCTCCCTGTGGGGGGC----- 2813
 QY 538 ProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyVal 557
 DB 2814 -----AACAAAGACAGAGAAACACCGCGGCTC 2840
 QY 558 ValLeuLysLeuProProAlaValAlaProValThrAlaSerTyrArgGlnGlnSerGln 577
 DB 2841 CTTGATGCTTCCCA----- 2855
 QY 578 SerAsnProLeuHisTyrMetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSer 597
 DB 2856 -----CCTTGGAGTCTCCAGACAGTTCCTGATTCCTGCGCCCAAGGCA----- 2900
 QY 598 LeuTyrProProArgGluAsnTyr-----IleArgGlyAla 609
 DB 2901 CTGGCCCAATCTGAGAGATTAACCTGTCATGATCATTTGTCGCTGCTCG 2951

Tue Apr 12 08:34:21 2005

us-09-920-705-3.rng

Page 24

Search completed: April 8, 2005, 09:07:22
Job time : 1324 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 08:37:33 ; Search time 370 Seconds

(without alignments)
3984.555 Million cell updates/sec

Title: US-09-920-705-3

Perfect score: 4801

Sequence: 1 MALSMKFRPADSGFQSNL.....LQIQKQKQSSGTPAGQP 901

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=xlh
-Q=/cgn2_1/uspto.epool_h/us09920705/runat_07042005_101814_20701/app_query.fasta_1.1095
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09920705_@CEN_1_1_75_@runat_07042005_101814_20701 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/prodata/1/ina/55_COMB.seq:*
2: /cgn2_6/prodata/1/ina/58_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCITUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	271.5	5.7	11231	4	US-09-949-016-14030
2	268	5.6	3192	4	US-09-949-016-2288
3	268	5.6	3228	4	US-09-949-016-2288
4	246.5	5.1	1477	4	US-09-976-594-794
5	246	5.1	2228	1	US-07-726-607C-1
6	246	5.1	2228	1	US-07-843-949A-1
7	246	5.1	2228	2	US-08-218-978-1
8	246	5.1	2356	4	US-09-949-016-4865
9	246	5.1	2356	4	US-09-949-016-4865
10	242.5	5.1	1401	4	US-07-843-949A-3
11	242.5	5.1	1401	2	US-08-218-978-3
12	236.5	4.9	8147	3	US-09-514-247A-9

13	215.5	4.5	1968	4	US-09-949-016-3886	Sequence 3886, Ap
14	208	4.3	4674	4	US-09-410-551B-25	Sequence 26, Ap1
15	208	4.3	4674	4	US-09-940-316B-25	Sequence 26, Ap1
16	207	4.3	1323	4	US-09-949-016-4868	Sequence 4868, Ap
17	207	4.3	1323	4	US-09-949-016-4868	Sequence 4869, Ap
18	201.5	4.2	4737	4	US-09-410-551B-30	Sequence 30, Ap1
19	201.5	4.2	4737	4	US-09-940-316B-30	Sequence 30, Ap1
20	200.5	4.2	1467	1	US-07-861-075-50	Sequence 50, Ap1
21	200.5	4.2	1467	1	US-08-120-827-50	Sequence 50, Ap1
22	200.5	4.2	1467	1	US-08-478-675-50	Sequence 50, Ap1
23	200.5	4.2	1467	1	US-08-461-027-1	Sequence 1, Ap11
24	200.5	4.2	1469	3	US-09-082-501-1	Sequence 1, Ap11
25	200.5	4.2	1469	3	PCT-US92-01836-1	Sequence 1, Ap11
26	200.5	4.2	2559	4	US-09-920-132-135	Sequence 135, Ap
27	200.5	4.2	2562	4	US-08-146-421-4	Sequence 4, Ap11
28	200.5	4.2	13987	2	US-08-804-227C-13	Sequence 13, Ap1
29	200.5	4.2	44377	2	US-08-804-227C-7	Sequence 7, Ap11
30	200.5	4.2	44377	2	US-08-804-198-1	Sequence 1, Ap11
31	200	4.2	2642	4	US-09-949-016-1282	Sequence 1282, Ap
32	199.5	4.2	2731	4	US-09-976-594-281	Sequence 281, Ap
33	198.5	4.1	9046	1	US-08-227-536-1	Sequence 1, Ap11
34	198.5	4.1	9046	5	PCT-US95-04662-1	Sequence 1, Ap11
35	195.5	4.1	1231	4	US-09-949-016-1766	Sequence 1766, Ap
36	195.5	4.1	1416	3	US-08-866-928B-2	Sequence 2, Ap11
37	195.5	4.1	1416	4	US-09-685-836-2	Sequence 2, Ap11
38	195.5	4.1	4767	4	US-09-410-551B-28	Sequence 28, Ap1
39	195.5	4.1	4767	4	US-09-940-316B-28	Sequence 28, Ap1
40	195.5	4.1	4818	4	US-09-410-551B-32	Sequence 32, Ap1
41	194.5	4.1	4818	4	US-09-940-316B-32	Sequence 32, Ap1
42	194.5	4.1	2277	1	US-08-676-967-5	Sequence 5, Ap11
43	194.5	4.1	2277	1	US-08-676-974-5	Sequence 5, Ap11
44	194.5	4.1	2277	2	US-09-098-487-5	Sequence 5, Ap11
45	194	4.0	4725	4	US-09-410-551B-24	Sequence 24, Ap1

ALIGNMENTS

RESULT 1
US-09-949-016-14030
Sequence 14030, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FASTSEQ for windows Version 4.0
SEQ ID NO 14030
LENGTH: 11231
TYPE: DNA
ORGANISM: Human
US-09-949-016-14030

Alignment Scores:
Pred. No.: 2.24e-11
Score: 271.50
Percent Similarity: 34.32%
Best local Similarity: 22.22%
Query Match: 5.66%
DB: 4
Gaps: 31

US-09-920-705-3 (1-901) x US-09-949-016-14030 (1-11231)
7 ProPhArg-----AlaApApSpScGlPhcGlnSerAmnLen 20

```

Db      3032 CTTTACAGAGAGTGAATGATTTCAACCGAGATATATACGACGATTAACCGAGCCTC 3091
Qy      21 TrpValGlySerLeuThrProGluThrThiGluSerAspLeuThrGluLeuPheGlyArg 40
Db      3092 TTCTTGGGCAACCTAGACATCATCTGTAACGAGAGTATTTAAGAAAGGCGTTGATCGC 3151
Qy      41 TyrGlyAspIleAspArgIleThrVal---TyrSerSerArgGly-----Phe 55
Db      3152 TTTCGACTCATACAGAGATGATATATCAAGAGCCTTCTCGCGGACGACTAGTACTTAC 3211
Qy      56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaIalysGluAlaLeuGlyArg 75
Db      3212 GGCCTTCTCAAAATTGAGAACTTAGATATGCTCACCGGCGCAAAATTAGCAATGCTCGC 3271
Qy      76 AlaAsnLeuAsnGlySerGlnIleValIleGluTyrAlaArgProAlaIysProCysLys 95
Db      3272 AAAATTATTAATTCGAAATCCTATCAAAATTGTTATGTAA---GCTACACCCACCACC 3328
Qy      96 SerLeuThrValGlyGlyIleGlyProAsnValSerLysAspAspLeuGluGluIuphe 115
Db      3329 CGCCTCTGGGTGGGAGGCTCGGACCTTGGGTTCTTCTGCTGCCCTGGACAGAAATTT 3388
Qy      116 SerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIle 135
Db      3389 GATCGATTGGCACCATACGACCATATAGACTACCGAAAGGTATGTGGCATATATC 3448
Qy      136 AspTyrTyrGluMetAspAspAlaLeuGlnAla---LysSerMetAsnGlyLysProMet 154
Db      3449 CAATATGAAGACCTGGATGACGCGATGCTGCTGACCCATATGGGGGCTTCCACCTT 3508
Qy      155 GlyGlyLys---SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysGlu 172
Db      3509 GGTGGCCAGATCGACGCTTAGAGTAGACTTTGCCGACACCGAAATGCTTACCGACGAG 3568
Qy      173 GlnTyr-----AlaGlySerTyrAspAsn 180
Db      3569 CAATATCTGACGCTCTGCCCTTGACTCATATAGACTGATGACAGATCTTTTGGACAT 3628
Qy      181 Arg-----AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyr 195
Db      3629 CGGGACACGACCCCTTTGAGGGGTCTCGGATAGACACCA---CCCTTACTATAC 3682
Qy      196 GluAspPheLysGlyAspValGlnProSerLysValLeuTyrIleGlyPheProProThr 215
Db      3683 AGAGATCGTATAGGACCTTATCTGACTCGAT---TGGGTGCCACCCCAACCCCA 3739
Qy      216 AlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyLysIle 235
Db      3740 GTCCCA---GACGCGACACTCGGACTGACGCTACTTGTGCTGCTTAC 3787
Qy      236 GluArgValLysSerTyrProSerArgAsnPheAla----- 247
Db      3788 GAGCCACTGATAGCTAGATGCGACGGCGGATGTGTCTTGGACCGGACAGAGT 3847
Qy      248 -----LeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGluIleu 264
Db      3848 GATCGAGATCTGCCAGACAGACAGACAGCAGCTTAGAAGCGAAGCTGCTGAGAGAGT 3907
Qy      265 GlnGlyArgLeuPhe-----AsnAsnProArgIleLysIleMetTyr 278
Db      3908 GGAGAGCGTCACTGGATAGTCTCTCTGAGAGTGAACCGCCAGCAAAAGTCACTGCCCT 3967
Qy      279 SerAspAspGluLeuProGluGlnAspAspThrSerPheTyrSerGlyMetLysArg 298
Db      3968 CCTTCTCTGACCGGACGTCCAGATTGAGC----- 3997
Qy      299 SerArgThrAspMetPheAsnAspAspProSerCysValSerSerProHisSerThrGly 318
Db      3998 AGTACCGGAGATGTTACACACGAC-----AATGATCGATCTTCCGT 4042
Qy      319 Ile-----ProGlySerMetLysArgProLeuArgGlyThrAsnGlnArgSerTyr 334

```

```

Db      4043 CTTCTTTGAAAGGCCCTCTCCAAATCAGACACAGACGAGTAGTTGGAGAAAGGC--- 4099
Qy      335 AsnGlyAlaGluTyrAsnAspValAlaGlyLysGluProAsnTyrArgProSerAla 354
Db      4099 ----- 4099
Qy      355 AsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGln 374
Db      4100 -----CAG 4102
Qy      375 GlyThrArgArgProMetArgSerAsnProAspSerTyrGluIleTyrAspProAlaGln 394
Db      4103 GGTGACACCGAGAC---CGTAAATCTGTGATCACTGAA--- 4141
Qy      395 LeuValArgGluSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMet 414
Db      4142 -----CGAGATAGAGAAAGCACCGACAACCTGCTCCCACTGAGGAGAAAGCCCTCTG 4192
Qy      415 GlyValAspGluArgSerPheGlyArgGly-----SerValAlaAlaArgProIle 431
Db      4193 AAAAAGAAACCGCTGTGATGGAGTGACCTAGACACACACATGCTTCTCCAAAGCTG 4252
Qy      432 ArgGlyPro-----ProAsp 436
Db      4253 AAGTCCCGTCCCAAGAAACAGATGGGGGACAGCCCTGTGGCATCAGCCTTCCCAA 4312
Qy      437 SerAspHisIleTyrArgGlyMet----- 444
Db      4313 CTCTGTTGGCTGGACGAGGATGCTTCTACTGAAACAGACAATTCTTCCAAACATG 4372
Qy      445 ---IleAlaLysGlyGlyThrProValCysValAlaArgCysVal-----ProMet 460
Db      4373 CATCTGTGACGGGTGACCTCCAAAGTGGCTGTAGTCTTCTTGGAGAGTTCAACTGGA 4432
Qy      461 GlyLysGlyIleGluThrLysLeuProGluValAlaAsnGlySerAlaArgThrAspLeu 480
Db      4433 GCGAAAGTGGCCAGCTCAAGATCATCAGCGTCCGTTTGAC---CAGCCCAAGTTG 4489
Qy      481 AsnMetLeuAlaLysHisTyrAlaValAla---IleGlyCysGluIleValPhePhe 498
Db      4490 GATGAAGTAACTGACAGCATCAAGATGACAGGCGCAATGATTCATCTTTGGCT 4549
Qy      499 ValProAspArgGluGlu-----AspPheAlaSer 508
Db      4550 GTGCTGGAAGTTCTGACAGCCGCTCTCTTCTCAGCTGATAGACACTGCCACT 4609
Qy      509 TyrThrGlu-----PheLeuArgTyrLeuSerSerLysAspArgAlaGly 523
Db      4610 TCTACTCAGAGGCCACTTAGAACCCTTGTGCTATTAAAGCAAAAGCAGGAGCCGGG 4669
Qy      524 ValAlaLysLeuAspAspArgIleThrThrLeuPheLeuValProProSerAspPheLeuThr 543
Db      4670 GTGATCAGCCTCCTGTGGGGGCG------ 4693
Qy      544 AspValIleuGlnValThrArgGlnLysArgLeuTyrGlyValIleuLysLeuProPro 563
Db      4694 -----AACAAAGCAAGAAACACCGGGTCTCTTACGCTTCCCA--- 4735
Qy      564 ProAlaValProValThrAlaSerTyrArgGlnLysSerLysAsnProLeuHisTyr 583
Db      4736 -----CCTGTGAGTTG 4747
Qy      584 MetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProAlaGlu 603
Db      4748 TCCAGCAGATTCCTGATTCCTCGCCGACAGCA-----CTGGCCAAATCTGAAGAA 4798
Qy      604 AsnTyr-----IleArgGlyAla 609
Db      4799 GATTACCTGTGATCATGATCATTCGTCGTGTGG 4831

```

RESULT 2

US-09-949-016-2288

/ Sequence 2288, Application US/09949016

```

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ. ID NOS: 207012
; SOFTWARE: PatSeq for Windows Version 4.0
; SEQ ID NO 2288
; LENGTH: 3192
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2288

Alignment Scores:
Pred. No.: 6,07e-12 Length: 3192
Score: 268.00 Matches: 155
Percent Similarity: 34.28% Conservative: 86
Best Local Similarity: 22.05% Mismatches: 248
Query Match: 5.58% Indels: 214
DB: 4 Gaps: 30

US-09-920-705-3 (1-901) x US-09-949-016-2288 (1-3192)

QY 7 ProPhaArg-----AlaAspAspSerGlyPheGlnSerAsnLeu 20
DB 1032 CCTTACAGAAAGCTGATGATGATTTCACTCCAGAGATGATCAGCAGCTAACCGAGCTC 1091
QY 21 TrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArg 40
DB 1092 TTCTTGGCACTTACGATCAGTCACTGACGAGAGATGATTTAAGAAAGGCGTTGATCGC 1151
QY 41 TyrGlyAspIleAspArgIleThrVal---TyrSerSerArgGly-----Phe 55
DB 1152 TTTCGAGTCATCAGAGATGATGATTCAGAGAGCGCTTCTCGCGCCAGACTAGTACTTAC 1211
QY 56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaGlyGluAlaLeuGlnGly 75
DB 1212 GCGTTCTCTCAAAATTGAGAACTTGAATGATGCTCACCGGCGCAAAATTGAGCAATGCTGCGC 1271
QY 76 AlaAsnLeuAsnGlySerGlnIleValIleGluTyrAlaArgProAlaLysProCysLys 95
DB 1272 AAAATTATATATTGGAAATCTTATCAAAATTTGTTATGTAA---GCTACACCCACCACC 1328
QY 96 SerLeuThrValGlyGlyIleGlyProAsnValSerLysAspAspLeuGluGluPhe 115
DB 1329 CGCTCTGGGGTGGAGGCGCTGGAGCCTTGGCTTCTGCTGCGCTGGCAGAGAAATT 1388
QY 116 SerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIle 135
DB 1389 GATCGATTGGCCACCTACGACCATAGTACTCCGAAAGGTGATGTTGGGCAATATATC 1448
QY 136 AspTyrTyrGluLeuAspAspAlaLeuGlnAla---LysSerMetAsnGlyLysProMet 154
DB 1449 CAGTATGAAGAGCTGGATGCGAGCGCATGCTGCTGGACCCATATGGGGGCTTCCCACTT 1508
QY 155 GlyGly-----SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysGlu 172
DB 1509 GGTGGCCAGATGAGCGCTTAGAGTAGACTTTGGCGACACGAAACATGTTTCCAGGAG 1568
QY 173 GlnTrp-----AlaGlySerTyrAspArg 180
DB 1569 CAGTATCTGACAGCGCTGCTGACCTATGATGAGCTGTGACAGATGCTTTTGGACAT 1628
QY 181 Arg-----AsnGlyAsnMetAsnIleLysArgProGlnTyrProHisSerTyr 195

```

```

DB 1629 CCGGCACACAGACCTTTGAGGGTCTCGGATAGGACACCA-----CCCTACTATATAC 1682
QY 196 GluAspPheLysGlyAspValGlnProSerLysValLeuTrpIleGlyPheProThr 215
DB 1683 AGAGATCGTATGAGGACCTTATCTGACTGTAT---TGGGTCCACCCCAACCCCA 1739
QY 216 AlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIle 235
DB 1740 GTCCGA-----GAGCGCAGACTCGGACTGACACTTCTGTGCTGCTTAC 1787
QY 236 GluArgValLysSerTyrProSerArgAsnPheAla----- 247
DB 1788 GAGCCACTGATAGCTTAATGCGAGCGGATGCTGTGCTTGGACCGGAGCAGAGGT 1847
QY 248 -----LeuValGluPheArgSerAlaGluGluAlaArgGlnCysLysGlyLeu 264
DB 1848 GATCGAGATCTGCCACGACAGACAGACAGCTTAGAGAAAGCTGCTGCTGAGAGAGT 1907
QY 265 GlnGlyArgLeuPhe-----AsnAsnProArgIleLysIleMetTyr 278
DB 1908 GAGAGAGTCATCTGGATAGTCTCTGAGAGTACCGCCCAAGAAAGCTACCTGCGCT 1967
QY 279 SerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArg 298
DB 1968 CCTTCTCCGACCGCAGTCCAGAAATTGAC----- 1997
QY 299 SerArgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGly 318
DB 1998 AGTAGCGGGATCGTTACAAACAGCGAC-----AATGATGATCTTCCCT 2042
QY 319 Ile-----ProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyr 334
DB 2043 CTTCCTTGAAGAGCGCTCTCCATACAGACAGACAGAGTATGTTGAGAAAGAC--- 2099
QY 335 AsnGlyAlaGluTyrAsnAspValAlaGlyLysGluProAsnTrpArgArgProSerAla 354
DB 2099 ----- 2099
QY 355 AsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGln 374
DB 2100 -----CAG 2102
QY 375 GlyThrArgArgProMetArgSerAsnProAspSerTrpGluGlyTyrAspProAlaGln 394
DB 2103 GGTGACAGACGAGAC---GTAAGAACTGTGATCAGCTGAA----- 2141
QY 395 LeuValArgGluSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMet 414
DB 2142 -----CGAATAGAAAGACCGGACAACTGCTCCCACTGAGGAAAGAAACCTCTG 2192
QY 415 GlyValAspGluArgSerPheGlyArgGly-----SerValAlaAlaArgProIle 431
DB 2193 AAAAAGAAAGACCGCTGTATGAGAGTGCACCTAGCAGACAGCACTCTTCTCAAGCTG 2252
QY 432 ArgGlyPro----- 436
DB 2253 AAGTCCCGCTCCAGAAACAGATGGGGGAGACGCCCTGTGGCATCAGCTTCTCCAA 2312
QY 437 SerAspHisIleTrpArgIleMet----- 444
DB 2313 CTCTGTTTGGCTGCGAGGAGCATGCTTCTTACATGAAAGACAACTTCTCTCAACATG 2372
QY 445 ---IleAlaLysGlyGlyThrProValCysValAlaArgCysVal-----ProMet 460
DB 2373 CATCTGTTCAGAGGTACCTCCAAAGTGGCTAGTACTTCTTGTGAGGGTTCAACTGGA 2432
QY 461 GlyLysGlyIleGluThrLysLeuProGluValAlaAsnCysSerAlaArgThrAspLeu 480
DB 2433 GGGAAAGTGGCCAGCTCAAGATCACTCAGCGCTCTCGTTTGGAC---CAGCCCAAGTTG 2489
QY 481 AsnMetLeuAlaLysHisTyrAlaValAla-----IleGlyCysGluIleValPhePhe 498

```



```

Db      2490 GATGAAGTAAGTCCAGCATCAAGTAGACAGGCGCCAAATGTTATGCATTTCTTTGGCT 2549
Qy      499 ValProasparglu-----AspPheAlaser 508
Db      2550 GTGGCTGGAACTTGACACGCGGTCTCTCTTCTCAGTCGATCAGACACGCGCCT 2609
Qy      509 TyrThrIgu-----PheLeuArgTyrIleuSerSerIysAspArgAlaGly 523
Db      2610 TCTACTCAGAGGCGCACTTAGAACCTTGTCCTATTAAAGCAAAAGCGAGCGCGG 2669
Qy      524 ValAlaIysIleuAspArgIYThrThrLeuPheLeuValProPheSerPheLeuThr 543
Db      2670 GTGATCAGCTCTCCCTGTGGGGGCGC-----2693
Qy      544 AspValIleuGlnValThrArgGlnIuArgLeuTyrGlyValIleuIysLeuProPro 563
Db      2694 -----AACAAAGACAGAGAAACACCGGGGTCTTCATGCGCTTCCCA-- 2735
Qy      564 ProAlaValProAlaThrAlaSerTyrArgGlnIuSerGlnSerIleuSerProLeuHleTyr 583
Db      2736 -----CCTTGTGAGTTC 2747
Qy      584 MetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgIu 603
Db      2748 TCCAGCAGTTCTGTGATTCCTGCGCAAGCA-----CTGGCCAAATCTGAAGAA 2798
Qy      604 AsnTyrIle 606
Db      2799 GATTACTG 2807

RESULT 3
US-09-799-451-752
Sequence 752, Application US/09799451
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhou, ping
APPLICANT: Goodrich, Ryle
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyan
APPLICANT: Zhang, Jie
APPLICANT: Xue, Aidong J.
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Ma, Yungting
APPLICANT: Yamazaki, Victoria
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Zhiwei
APPLICANT: Wang, Dunrui
APPLICANT: Yang, Yonghong
APPLICANT: Wehrman, Tom
APPLICANT: Ghosh, Reena
APPLICANT: Dimanac, Radoje T.
TITLE OF INVENTION: No. 6783969el Nucleic Acids and
FILE REFERENCE: 803
CURRENT APPLICATION NUMBER: US/09/799,451
CURRENT FILING DATE: 2001-03-05
SOFTWARE: pt_FL_genes Version 2.0
SEQ ID NO 752
LENGTH: 3228
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (108) ..(2882)
US-09-799-451-752

Alignment Scores:
Pred. No.: 6,18e-12 Length: 3228
Score: 268.00 Matches: 155
Percent Similarity: 34.28% Conservative: 86

```

```

Best Local Similarity: 22.05% Mismatches: 248
Query Match: 5.58% Indels: 214
DB: 4 Gaps: 30

US-09-920-705-3 (1-901) x US-09-799-451-752 (1-3228)

Qy      7 ProPheArg-----AlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
Db      1044 CTTTCAGAGAGATGATGATGATGATTCACCCGAGATGATCAGCGATCAACCGAGCTC 1103
Qy      21 TrpValGlySerLeuThrProGluThrThrGlnIuSerAspIleuThrGlnLeuPheGlyArg 40
Db      1104 TTTCTGGGCAACCTTGACATCACTTAACGGAGAGATGATTAAGAGGCGCTTGATGCG 1163
Qy      41 TyrGlyAspIleAspArgIleThrVal--TyrSerSerArgIly-----Phe 55
Db      1164 TTTGAGATCATCAGAGAGATGATTAAGAGGCGCTTCGCGGCACAGTATGATCTTAC 1223
Qy      56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaIysGluAlaLeuGlnIly 75
Db      1224 GCGTTCTCAATTGAGAACTTATGATATGTCACCGCGGCAATTAGCAATGCTGGC 1283
Qy      76 AlaAsnLeuAsnGlySerGlnIleTyrIleGluTyrAlaArgProAlaIysProCyls 95
Db      1284 AAATTTATATTCGGAATCTTCAATTAATTTGTTATGTA--GCAACCCACACACC 1340
Qy      96 SerLeuTrpValGlyIlyIleGlyProAsnValSerIysAspAspLeuGluGluPhe 115
Db      1341 CGCCTCTGGGTGGAGGCGCTGGACCTTGCGTTCCTTCTGCGCTCGCACAGAAATT 1400
Qy      116 SerIysPheGlyIlyIleGluAspPheArgPheLeuArgIuArgIystrAlaPheIle 135
Db      1401 GATCATTTGGACACCATTCGACCATTAACCGAAAGGTATGTTGGCGATATATC 1460
Qy      136 AspTyrTyrGluMetAspAspAlaLeuGlnAla--LysSerMetAsnGlyLysProMet 154
Db      1461 CAGATGAAGACCTTGATGACCGCATGCGCTCGGACCATATGCGGGGCTTCCACTT 1520
Qy      155 GlyGlyI-----SerPheLeuArgValAspPheLeuArgSerGlnAlaProLysIysGlu 172
Db      1521 GGTGGCCAGATGACCGCTTGAAGTACTTCCGACACCGAACATCTTACACAGAG 1580
Qy      173 GlnTrp-----AlaGlySerTyrAspAsn 180
Db      1581 CAGTATTCGACGCTCGCTTACCTTATGATGAGCTGTGACAGATGCTTTTGACAT 1640
Qy      181 Arg-----AsnGlyAsnMetAsnHleLysAspProGlnTyrProHisSerTyr 195
Db      1641 CGGGCACACAGACCTTTGAGGGGTGCTGGGATATGACACCA-----CCCTTACTATAC 1694
Qy      196 GluAspPheIysGlyAspValGlnProSerIysValLeuTrpIleGlyPheProProThr 215
Db      1695 AGAGATCGTATAGGAGCTTATCTGACTGTAT--TGGGTGCCACCCACCCCA 1751
Qy      216 AlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyIuIle 235
Db      1752 GTCCGA-----GAACGAGACATCGGACTGACACTTCTGTGCTGCTTAC 1799
Qy      236 GluArgValIysSerTyrProSerArgAsnPheAla-----247
Db      1800 GAGCACTGATGATGCTTATGATTCGACGCGGATGTGTGCTTGGACCGGACAGAGGT 1859
Qy      248 -----LeuValGluPheArgSerAlaGluGluAlaArgGlnCylsIysGluIleu 264
Db      1860 GATCGAGATCTGCCACAGACAGACAGACACCTTAGAGAGAGAGGCTGCTGAGAGAGT 1919
Qy      265 GlnGlyArgLeuPhe-----AsnAsnProArgIleIysIleMetTyr 278
Db      1920 GAGAGACGTATCTGATAGTCTCTCTGAGAGTACCGCCACAGAAAGTACTACGCGCT 1979
Qy      279 SerAsnAspGluLeuProProGluGlnAspPheThrSerPheTyrSerGlyMetLysArg 298
Db      1980 CTTTCTCTGACCGCAGTCCAGAAATTGAC-----2009

```

QY	299	SeKpRgHtHArMeRcPhdAspRnsAspProSerCyValSerSerProHISerThGly	318
Db	2010	AGTAgCCGGGAGTCGTTACACAGCCAC-----AAATGATCATCTTCCCGT	2054
QY	319	Ile-----ProGlySerMetArpProLeuArgGlyThrAsnGluArpSerTyr	334
Db	2055	CTTCTCTTGAAGAACCCCTCTCCATCATGACAGACAGAGGTAGTTGGAGAGGC---	2111
QY	335	AsnGlyLaGluTyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAla	354
Db	2111	-----	2111
QY	355	AsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGln	374
Db	2112	-----CAG	2114
QY	375	GlyThrArgArgProMetArpSerAsnProAspSerTrpGluGlyTyrAspProAlaGln	394
Db	2115	GGTGCAGACGACAGAC---CGTAAAACTCTGCATCATGCTGA---2153	
QY	395	LeuValArgGluSerLysArgThrArgTrpArgTrpArgLysValAspArgLysPheThrProMet	414
Db	2154	-----CGAGATAGGAGACACCGGACAACTGCTCCCACTAGAGGAAAAAGCCCTGTG	2204
QY	415	GlyValAspArgLysArgSerPheGlyArgGly-----SerValAlaAlaArgProIle	431
Db	2205	AAAAAGAAAGAACCGCTGATGGAGAGTGACACTGACACACACAGCTCTCTCCAAAGTG	2264
QY	432	ArgGlyPro-----ProAsp	436
Db	2265	AAAGTCCCGGTCCCAAAACAGATGGGGGACAGCCCTGTGGCATGACCTTCCCAA	2324
QY	437	SerAspHisIleTrpArgLysMet-----	444
Db	2325	CTGCTGTTGGCGCTGSCAGGCGATGCTTTACTAGAAACACACAACTTCTTCCAAAGTG	2384
QY	445	--IleAlaLysGlyLysThrProValCysCysValArgCysVal-----ProMet	460
Db	2385	CATCTGTTGACAGGGTACCTCCAAAGTGGCTAGTACTTCTTGGAGAGGTTCAACTGGA	2444
QY	461	GlyLysGlyIleGluThrLysLeuProIleValValAsnCysSerAlaArgThrAspLeu	480
Db	2445	GGCAAAAGTGGCCCACTACAGTACCTGAGGCTCTCGTTGGAC---CACCCCAAGTTG	2501
QY	481	AsnMetLeuAlaLysIleTyrAlaValAla-----IleGlyCysGluIleValPhePhe	498
Db	2502	GATGAAGTAACTCCAGCAGCATAAATGACAGGGCCCAATGGTTATGCCATTTCTTTGGCT	2561
QY	499	ValProAspArgGluGlu-----AspPheAlaSer	508
Db	2562	GTGCGCTGGAAGTTGTGACAGCCCGCTCTCTCTCTCCAGCTGATCAGACACTGCCACT	2621
QY	509	TyrThrGlu-----PheLeuArgTyrLysSerSerLysAspArgAlaGly	523
Db	2622	TCTACTCAGAGGCCACTTGAAGAACTTGTGCTCTATTAAAGCAAAAGCAGCGCGGG	2681
QY	524	ValAlaLysLeuAspArgLysThrThrLeuPheLeuValProProSerAspPheLeuThr	543
Db	2682	GTGATCAGCCCTCCCTGTGGGGGGG---	2705
QY	544	AspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValIleLeuLysLeuProPro	563
Db	2706	-----AACAAAGACAAAGAAACACCGGGGGCTTTCATGCTTCCCA--	2747
QY	564	ProAlaValProValThrAlaSerTyrArgGlnLysSerGlnSerAsnProLeuHisIleTyr	583
Db	2748	-----CTTGGAGAGTTC	2759
QY	584	MetAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGlu	603
Db	2760	TCCAGCAGGTCTCGGATTTCCCTGCGCAAGGA-----CTGGCCAAATCTGAAGA	2810

```

QY      604 Aantyrile 606          :::|:::
Db      2811 GATTACCTG 2819
RESULT 4
US-09-976-594-794
; Sequence 794, Application US/0976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 60/240,409
PRIORITY FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 1143
SOFTWARE: PERL Program
SEQ ID NO 794
LENGTH: 1477
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6673549 4436028CB1
US-09-976-594-794
Alignment Scores:
Pred. No.: 1,096-10 Length: 1477
Score: 246.50 Matches: 99
Percent Similarity: 39.53% Conservative: 69
Best Local Similarity: 23.29% Mismatches: 172
Query Match: 5.13% Indels: 85
DB: 4 Gaps: 21
US-09-920-705-3 (1-901) x US-09-976-594-794 (1-1477)
QY      4 SerMetIysPPropheArgAlaAspAspSerGIyPhgInserAsnAsnLeuTrpValGIy 23
Db      74 TCCTCGCAGCCACCACCATGATGTGAAGAAGACGG--CAGCCCCGCACTTACTAGTAGGT 130
QY      24 SerLeuthrProGluThrThrGlusSerAspleuThrgluleuPhgeIyArgTyGIy--- 42
Db      131 AACCTTCAGAGAGTGTCACAGAAAGTCCTTACTTACTTCAGTTGTTTCAGTCAGATTGGACCC 190
QY      43 -----AspieleApArgIIeThrValIySerSerArgGIy---PheAlaPheIle--- 58
Db      191 TGTAAGAGCTGTAAATGATATGATACAGAGCATVCAAGCAATGACCCATATTGCTTGTGGAA 250
QY      59 TyTrYrArghISvalGIgluaIalavalalaIylsgualaleugInGIyAlaAsndeu 78
Db      251 TTTTATGAAAC--AGAGATGCAGCTGCTGCATTAGCTGTATGAAATGGAGAAAATT 307
QY      79 AsnGIySerGlnIleIyEllegIutryla--ArgProAlaIyProcylAspSer--- 96
Db      308 TTGGGAAAGAGGTCAAAGTAACTGGGCAACACACCAAGTAGCCAGAAAAAGATACT 367
QY      97 -----LeuTrpValIGlyGIyIleGIyProAsnValSetIyAspAspIeu 111
Db      368 TCCATACATTCATCGATGTGTGTGGGATTTGAGTCCAAGAAATTAACAAGAAATATC 427
QY      112 GluGIgluGIuPheserIyPheGIyIyEllegIuAspPheArgPheIeuArgIuArgIyS 131
Db      428 AAATCAGCATTTGCCCTTTGGTAAATATCGAATGCCCGGATAGTTAAAGACATGGCA 487
QY      132 Thr-----AlaPheIleAspTyTrGIuMeIAspAspAlaIeuGI 145
Db      488 ACTGGAATAATCCAAAGGCTATGCTTTGTATCTTTTATATACAAACTGATGACGAATAT 547
QY      146 Ala---LyseRmetAenGIyLysPromeGelIyIySerPheIeuArgValAspPheIeu 164
Db      548 GCAGATTGTGATATGGGGGTCAAGTGTGGGTGGTGGTGTGGTCAATTCGAACCAATTTGGCC 607

```

```

QY 165 -----ArgserGlnAlaProLysLeuGluGlnTrpAlaGlySerTyrAspAsnArgAsn 162
Db 608 ACTGCTAAACCACTGACCTTAAAGACAACT-----GAAACCAACACT 652
QY 183 GlysMetAsnHisLysProGlnTyrProHisSerTyrGluPhePheLysGlyAspVal 202
Db 653 AAGCAGTTGAGA-----TTGAAAGTGTAGTAAACCACTCA 668
QY 203 GlnProSerLys-----ValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsn 220
Db 689 AGTCAAAAATTTGACTGTGTACTGTGAGAGATTGCTGTGGTTAACA----- 729
QY 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGluIleGluValGlyValSer 240
Db 740 ---GATCAGCTTATGAGACACAGACTTTCACCATTTTGACAAATTATGAAATTAAGGTT 796
QY 241 TyrProSerArgPheAsnPheAlaLeuValGluPheArgSerAlaGluGluValArgGlnCys 260
Db 797 TTCCAGAAAAGGGCTTATTCATTTGTGAGATTTCAGATTTCACATGAAAGTGCAGCCCATGCC 856
QY 261 LysGluGluLysLeuGlnArgLysLeuPheAsnAsnProArgLysIleLysIleMetTyrSerAsn 280
Db 857 ATTGTTTCGGTGAACGGTACTACAGATTGAAGACATGTGGTTAAA---TGCTATTGGGGT 913
QY 281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
Db 914 AAGAATCTCTCT-----GATATGACTTAAACTTCCACAGGTTGACTATGACTCA 964
QY 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
Db 965 TGG----- 967
QY 321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsn 340
Db 968 GGCCANTGAGCCCAAGTATGCA-----AACCCACAACAGTATGGA 1009
QY 341 AspValValGlyLysGluProAsnTrpArgArgPro----- 352
Db 1010 CAGTATATGCAAAAT-----GGGTGGCAAGTACCGCTTATGAGATATACGGGCAACA 1063
QY 353 ---SerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro-----GlyIle 368
Db 1064 TGAATACCAACAGAGATTGAGATGATCAATCACTTCGCTGCTGATGATGGATGTTT 1123
QY 369 LeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGlu 388
Db 1124 GGTGCTCAGCTCCCAAGGACAGCTCCTCCCTGTAATACCTCTCTTAACCAAGCC 1183
QY 389 GlyTyrAspProAla 393
Db 1184 GGATATGATATGGA 1198

```

RESULT 5

US-07-726-607C-1

Sequence 1, Application US/07726607C

Patent No. 5288407

GENERAL INFORMATION:

APPLICANT: Anderson, Paul J.

APPLICANT: Streuli, Michel

APPLICANT: Schloegman, Stuart F.

TITLE OF INVENTION: DNA ENCODING A PROTEIN

TITLE OF INVENTION: ACTIVE IN LYMPHOCYTE

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSER: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110-2804

COMPUTER READABLE FORM:

```

MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/726,607C
FILING DATE: July 10, 1991
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 07/460,678
FILING DATE: January 5, 1990
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/021002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2228
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-07-726-607C-1
Alignment Scores:
Pred. No.: 2,286-10
Score: 246.00
Percent Similarity: 39.65%
Best Local Similarity: 21.20%
Query Match: 5,128
DB: 1
Gaps: 17
US-09-920-705-3 (1-901) x US-07-726-607C-1 (1-2228)
QY 20 LeuTrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGly 39
Db 81 CTATACGTCGGTAAACCTTCCAGAGATGTGACAGAACTCTTAATTGCAACTCTTTAGC 140
QY 40 ArgTyrGlyAspIleAspArgIle---ThrValTyrSerSerArgIly-----Phe 55
Db 141 CAGATTGGACCTGTAAATACTGCAAAATGATTATGATAGCTGAAATGATCCCTAT 200
QY 56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGly 75
Db 201 TGTTTTGTGAGTTTCATGAGCATGTCATGACAGCTGCACATTAGCTGCTATGATGGA 260
QY 76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAla---ArgProAlaLysProCys 94
Db 261 CGGAAGATATAGGTGTAAGAGATGAAAGTGAATTGGGCAACACCTTACAGCTCAAAAG 320
QY 95 LysSer-----LeuTrpValGlyLysIleGlyProAsnValSerLys 108
Db 321 AAAGATACAAAGCATCATTTCCATGCTCTTGTGTCATCTCAGCCAGAAATTACAACT 380
QY 109 AspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArg 128
Db 381 GAAGATATAAAGCTGCTTTTGACCACTTGGAGAAATATCAATGCCAGTGGTAAA 440
QY 129 GlnArgLysThr-----AlaPheIleAspTyrTyrGluMetAspAsp 142
Db 441 GACATGCAACAGGAAGTCTAAGGATAGCTGCTTGTCTCTTTTCAACAAATGAGAT 500
QY 143 AlaLeuGlnAla---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgVal 161
Db 501 GCTGAAAACCCCATTTCAACAGATGGGTGGCCAGTGGCTGTGTGGAAGACAAATCAGAACT 560
QY 162 AspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySerTyrAspAsnArg 181
Db 561 AACTGGGCAACCCGAAAGCTCCCGCTCCAAAG-----AGTACATATGATGATCAAT 611

```

```

Oy 182 AengIyAenMetAenHnIeLyeBProGIntYrProHIsSerTYrGluAspPheLyeGlyAsp 201
Db 612 ACCAAACAGCTA-----TCATATGAGAGCTTTGAATAG 647
Oy 202 ValGInProSerLys---ValLeuTriLeGlyPheProToThAlaThrGlnCysAsn 220
Db 648 TCTAGTCCAGCAAGACTGATACGTATACGTGGA-----GGGTACTCTTGGGCTA 698
Oy 221 AspGluGlnIleLeuHIsAsnAlaMetIleLeuPheGlyGluIleGluValIlySer 240
Db 699 ACAGAAACAATGATGCGTCAGACTTTTCCACCATTTGGACAATGATGAATTCGAGTTC 758
Oy 241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
Db 759 TTTCAGATTAAGATATTCATTCTTTGTGCTTCAATTCCTCAATGAAGTGCACACATGCA 818
Oy 261 LysGluGlnIleLeuGlnGlyArgLeuPheAsnAsnProAlaGlyIleLysIleMetYrSerAsn 280
Db 819 ATTGTTTCTGTTAAAGTACTACTACCTTGAAGTCACTGTTGGAAATGCTATTGGGCGAA 878
Oy 281 AspGluLeuProProGluGlnAspThrSerPheTYrSerGlyMetLysArgSerArg 300
Db 879 GAAACTCTT----- 887
Oy 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHIsSerThrGlyIlePro 320
Db 888 ---GATATGATTAATCC-----GTGCAACAGCAGAAATCAAAATTGGATATTC 932
Oy 320 ----- 320
Db 933 CAACCTTATGGCCAGTGGGCGCCAGTGTATGGAATGCACAACAATGGCCAGTATATG 992
Oy 321 ----glySerMetArgProLeuArgGlyThrAsnGluArgSerTYrAsnGlyAlaGlu 338
Db 993 CCTAATGCTTGCAAGATTCCTGCATATGGAATGATATGGCCAGGATGGAACAGACAGA 1052
Oy 339 TyrAsnAspValIleGlyLysGluProAsnTriPArgProSerAlaAsnGlyThryGly 358
Db 1053 TTTATATGACACACACTCTTCTGCACCA---TGATGGAGCACAAAAT-----TATGGA 1100
Oy 359 IleLeuProSerProThrGlyPro-----GlyIleLeuProSerProAlaGlnGlyThr 376
Db 1101 GTGCACCG---CTCAAGGGCAAAATGGCAGACATGTTGCCCAATCAGCCTTGGGAT 1157
Oy 377 Arg 377
Db 1158 CGA 1160

RESULT 6
US-07-843-949A-1
: Sequence 1, Application US/07843949A
: Patent No. 5340935
: GENERAL INFORMATION:
: APPLICANT: Anderson, Paul J.
: APPLICANT: Screllin, Michel
: APPLICANT: Schlossman, Stuart F.
: TITLE OF INVENTION: DNAS ENCODING PROTEINS ACTIVE IN
: TITLE OF INVENTION: LYMPHOCYTE-MEDIATED CYTOTOXICITY
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: COMPUTER: IBM PS/2 Model 502 or 55SX
: OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
: SOFTWARE: WordPerfect (Version 5.0)
: CURRENT APPLICATION DATA: US/07/843,949A
: APPLICATION NUMBER: US/07/843,949A

```

```

1      FILING DATE: 19920219
2      CLASSIFICATION: 435
3      PRIOR APPLICATION DATA:
4      APPLICATION NUMBER: 07/460,678
5      FILING DATE: January 5, 1990
6      APPLICATION NUMBER: 07/726,607
7      FILING DATE: July 10, 1991
8      ATTORNEY/AGENT INFORMATION:
9      NAME: Frazer, Janis K.
10     REGISTRATION NUMBER: 34,819
11     REFERENCE/DOCKET NUMBER: 00530/0210030
12     TELECOMMUNICATION INFORMATION:
13     TELEPHONE: (617) 542-5070
14     TELEFAX: (617) 542-8906
15     TELEX: 200154
16     INFORMATION FOR SEQ ID NO: 1:
17     SEQUENCE CHARACTERISTICS:
18     LENGTH: 2228
19     TYPE: nucleic acid
20     STRANDEDNESS: double
21     TOPOLOGY: linear
22     US-07-843-949A-1

```

Alignment Scores:

Pred. No.:	2,28e-10	Length:	2222
Score:	246.00	Matches:	85
Percent Similarity:	39.65%	Conservative:	74
Best Local Similarity:	21.20%	Mismatches:	158
Query Match:	5.12%	Indels:	84
DB:	1	Gaps:	17

US-09-920-705-3 (1-901) X US-07-843-949A-1 (1-2228)

Oy 20 LeuTrpValGIySerIeuThrProclunThrhGluSerAaPleuThnGluLeuPhcGIy 39
 Db 81 CTATACGTGGTAACCTTTCACAGATATGACAGAAAGCTTAATTCGTCAACCTCTTACG 140
 Oy 40 ArgTyGIyAaPleAaPrlleAaPrlle---ThrValTySerSerAaGIy-----Phe 55
 Db 141 CAATTTGGACCTTTGTAAAACCTGCAAATATGATATGATACAGCTGCAAAATGATCCCTAT 200
 Oy 56 AlaPheIIeTyTrTyTrArgHieValGluGluValAlaAlaIySGuIAlaLeuGInGly 75
 Db 201 TGTATTGTGAGATTTCATGACATGCTCATGCACGCTGCAGCATTAAGCTGCTATGAATGSA 260
 Oy 76 AlaAsnLeuAaSnGlySerGlnIleIySileGluTyAla---ArgProAlaIyProCys 94
 Db 261 CGAGAGATATATGGTATGAGAGTCAAAATGTAATGGGGACAAACCCCTTACAGTCAAAAG 320
 Oy 95 LysSer-----LeuTrpValGIyGlyIleGlyProAenValSerIys 108
 Db 321 AAAGATATCAAGCAATCATTTCCATGCTTTTGTGTATCTTCAGCCAGAAATTACAACT 380
 Oy 109 AspAsPleuGluGluGluIuPheSerIySphcGIySIIeGluAaPheArgPheLeuArg 128
 Db 381 GAAGATATATAAAGCTCTTTTGCACCATTTGGAAATAATCAGATGCCCGAGTGTAAAA 440
 Oy 129 GluArgLysThr-----AlaPheIIeAaPrlTyTyrgIuMeAaAsp 142
 Db 441 GACATATGGCAACGGAAAGTCTAAGGATATATGGCTTTGTCTCTTTTCAACAAATGGGAT 500
 Oy 143 AlaLeuGluAla---LysSerMeAaSnGlyIySPrometGIyIySerPheLeuArgVal 161
 Db 501 GCTGAAAAGCCCATTTCAACAGATGGGTGGCCAGTGGCTGTGGAAAGCAATCAGAACT 560
 Oy 162 AspPheLeuArgSerGlnAlaProIySlySGluGlnTrpAlaGIySerTyraaPaaArg 181
 Db 561 AACTGGGACCAACCGAAAGCTCCCGCTCCAAG-----AGTACATATGAGTCAAAAT 611
 Oy 182 AaSnGlyAaSnMeCaSnHieIySbProclunTyProHieSerTyrgIuAaPheIySbGIyAaP 201
 Db 612 ACCAAACAGCTA-----TCATATGATATGAGTTGTAATACG 647

```

Qy 202 ValGlnProSerIys---ValLeuTrpIleGlyPheProProThraIaThrGlnCysAsn 220
Db 648 TCTAGTCCAGCAACTGACTGATCTGATCTGCA-----GGTGTACTCTCGGGCTA 698
Qy 221 AspGluInIleLeuHsAenAlMetIleLeuPheGlyIleGluArgValIysSer 240
Db 699 ACAGAACCAACTAAAGCCGTCAGACTTTTTCACCATTTGGACAAATATGAAATTCGAGTCC 758
Qy 241 TyrProSerArgAsnAlaLeuValGluPheArgSerAlaGluAlaArgGlnCys 260
Db 759 TTTCCAGATTAAGATATTCATTGTTGCGTCAATTCCTCAATGAAAGTCAGACATGCA 818
Qy 261 LysGluGlyLeuGlnGlyArgLeuPheAsnProArgIleLysIleMetIlySerAsn 280
Db 819 ATTGTTCTGTTATGATGATACCATGTAAGTCATGTTGAAATCTTATGGGGCAAA 878
Qy 281 AspGluLeuProProGluGlnAspThrSerPheTyrSerGlyMetIlySarSerArg 300
Db 879 GAACCTCTT----- 887
Qy 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHISerThrGlyIlePro 320
Db 888 ---GATATGATTAATATCC-----GTGCACACAGACAAATGATGATATATCC 932
Qy 320 ----- 320
Db 933 CAACCTTATGCGCAGTGGCGCAGTGGTATGTAATGCAACAAATTTGGCCAGTATATG 992
Qy 321 -----GlySerMetArgProLeuArgGlyThrAsnGluArgSerIlyrAsnGlyAlaGlu 338
Db 993 CCAATATGTTGGCAAGTCTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 1052
Qy 339 TyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGly 358
Db 1053 TTTAATCAGACACAGCTTCTGACACA---TGATGGACCAAT-----TATGCA 1100
Qy 359 IleLeuProSerProThrGlyPro-----GlyIleLeuProSerProAlaGlnGlyThr 376
Db 1101 GTTCAACCG---CCTCAAGGGCAAAATGCGACATGTTGCCCAATCAGCTTCTGGGTAT 1157
Qy 377 Arg 377
Db 1158 CGA 1160

```

RESULT 7
 US-08-218-978-1
 ; Sequence 1, Application US/08218978
 ; Patent No. 5837811
 ; GENERAL INFORMATION:
 ; APPLICANT: Anderson, Paul J.
 ; APPLICANT: Streuli, Michel
 ; APPLICANT: Schlossman, Stuart F.
 ; TITLE OF INVENTION: DNAS ENCODING PROTEINS ACTIVE IN
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: Fish & Richardson
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; COMPUTER: IBM PS/2 Model 502 or 55SX
 ; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
 ; SOFTWARE: WordPerfect (Version 5.0)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/218,978
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/843,949

```

; FILING DATE: February 19, 1992
; APPLICATION NUMBER: 07/460,678
; FILING DATE: January 5, 1990
; APPLICATION NUMBER: 07/726,607
; FILING DATE: July 10, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/021003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2228
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-218-978-1
;
; Alignment Scores:
; Pred. No.: 2,28e-10 Length: 2228
; Score: 246.00 Matches: 85
; Percent Similarity: 39.65% Conservative: 74
; Best Local Similarity: 21.20% Mismatches: 158
; Query Match: 5.12% Gaps: 84
; DB: 2 Indels: 17
;
; US-09-920-705-3 (1-901) x US-08-218-978-1 (1-2228)
;
Qy 20 LeuTrpValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGly 39
Db 81 CTATGCTGCTGATTAATCTTCCAGAGATGTCAGAGATCTTAATCTGCAACTCTTTAGC 140
Qy 40 ArgTyrGlyAspIleAspArgIle---ThrValTyrSerSerArgGly-----Phe 55
Db 141 CAGATTGGACCTTGAATAAATCGCAAAATGATTTATGATATGATGATGATGATGATGATGAT 200
Qy 56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGly 75
Db 201 TGTTTTGTGAGATTTCATGACATGTCATGCTGACCTGACATGATTAAGCTGATGATGA 260
Qy 76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTrpAla---ArgProAlaLysProCys 94
Db 261 CGGAAGATATGCGTGAAGAGTCAAGATGATGGCAACACCCCTTAGCAGTCAAAAG 320
Qy 95 LysSer-----LeuTrpValGlyGlyIleGlyProAsnValSerIys 108
Db 321 AAAGATACAAAGCATCATTTCCATGCTCTTTGTTGTTGATGATCTAGCCAGAAATTAACA 380
Qy 109 AspAspLeuGluGluGluPheSerLysPheGlyLysIleGluPheAlaPheLeuArg 128
Db 381 GAAGATATTAAGAGCTGCTTTTGACCACTTTGGAAGATATCAAGATGCGGAGTGA 440
Qy 129 GluArgLysThr-----AlaPheIleAspTyrTyrGluMetAspAsp 142
Db 441 GACATGGCAACAGAAAGTCAAGGATATGCGTTTCTCTCTTTTCAACAAATGGAT 500
Qy 143 AlaLeuGlnAla---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgVal 161
Db 501 GCTGAACCGCATTTCAACAGATGGTGGCAGTGGCTTGATGAAGCAAAATCAAGACT 560
Qy 162 AspPheLeuArgSerGlnAlaProLysValGluGlnTrpAlaGlySerTyrAspAsnArg 181
Db 561 AACTGGGCAACCGGAAGGCTCCCGCTCCAAAG-----AGTCAATATGATGATCA 611
Qy 182 AsnGlyAsnMetAsnHisLysProGluTyrProHisSerTyrGluAspPheLysGlyAsp 201
Db 612 ACCAAACAGCTA-----TCATATGATGATGATGATGATGATGATGATGATGATGAT 647
Qy 202 ValGlnProSerIys---ValLeuTrpIleGlyPheProProThraIaThrGlnCysAsn 220

```



```

Db      1122 CCTAATGTTGGCAAGTTCCTGCATATGGAATGATGCGCAGCATGAAACCGAAGGA 1181
Qy      339 TyrAsnAspValValGlyLysGluProAsnTPrArgArgProSerAlaAsnGlyThrGly 358
Db      1182 TTTATCAAGACACAGCTCTTGCACCA---TGATGGGACCAAAAT-----TANGA 1229
Qy      359 IleLeuProSerProThrGlyPro-----GlyIleLeuProSerProAlaGlyGlyThr 376
Db      1230 GTGCACCG---CCTCAAGGGCAAAATGAGCAGCATGTGGCCCAATCAGCCTTCTGGGTAT 1286
Qy      377 Arg 377
Db      1287 CGA 1289

RESULT 9
US-09-949-016-4866
; Sequence 4866, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4866
; LENGTH: 2356
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4866

Alignment Scores:
Pred. No.: 2,486-10 Length: 2356
Score: 246.00 Matches: 85
Percent Similarity: 39.65% Conservative: 74
Best Local Similarity: 21.20% Mismatches: 158
Query Match: 5.12% Indels: 84
DB: 4 Gaps: 17

US-09-920-705-3 (1-901) x US-09-949-016-4866 (1-2356)
Qy      20 LeuTPrValGlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGly 39
Db      210 CTATACGTCGGTAACCTTTCCAGAGATGACGAAAGCTCTAATTCGTGCACTCTTTAGC 269
Qy      40 ArgTyrGlyAspIleAspArgIle---ThrValTyrSerSerArgGly-----Phe 55
Db      270 CAGATTGACCTTGTAATAAACTGCAAAATGATTATGATACAGCTGGAATGATCCCTAT 329
Qy      56 AlaPheIleTyrTyrArgHisValGluGluAlaValAlaIleLysGluAlaLeuGlnGly 75
Db      330 TGTGTTGTCGAGTTTCATACAGATGTCATGCACTGCGCACTTACGCTATGATGATGA 389
Qy      76 AlaAsnLeuAsnGlySerGlnIleLysIleGluTyrAla---ArgProAlaLysProCys 94
Db      390 CCGAAGATATATGCGTAAAGAGTCAAGTGAATGGGCAACAACCCCTTAGCAGTAAAG 449
Qy      95 LysSer-----LeuTPrValGlyGlyIleGlyProAsnValSerLys 108
Db      450 AAAGTATACAGCAATCATTCATGCTCTTGTGTGATCTCAAGCCGCAAAATTAACAAC 509
Qy      109 AspAspLeuGluGluLysPheSerLysPheGlyLysIleGluAspPheArgPheLeuArg 128
Db      510 GAAATATATAAAACGCTTTTGCACACATTGGAAGAAATATACAGATGCCGAGTGATAAA 569

```

```

Qy      129 GluArgLysThr-----AlaPheIleAspTyrTyrGluMetAspArg 142
Db      570 GACATGCGCAACGAAAGCTAAGGATATATGCTTTGCTCTTTTCAACAAATGCGAT 629
Qy      143 AlaLeuGlnAla---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgVal 161
Db      630 GCTGAAMACGCTATTCACAGATGGGTGCGCAGTGGCTTGTGGGAAGCAAAATCAGAACT 689
Qy      162 AspPheLeuArgSerGlnAlaProLysGluGlnTPrAlaGlySerTyrAspAsnArg 181
Db      690 AACTGGGCAACCCGAAAGCCTCCCTCCAAAG-----AGTACATATGAGTCAAT 740
Qy      182 AsnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAsp 201
Db      741 ACCAAACAGCTA-----TCATATGATGAGTGTAAATCAG 776
Qy      202 ValGlnProSerLys---ValLeuTPrIleGlyPheProProThrAlaThrGlnCysAsn 220
Db      777 TCTAGTCCAGCAAGCACTGACTGATATACGTGGA-----GGTGTACTTCTGGGCTA 827
Qy      221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
Db      828 ACAGAACACTATATGCGTCAGACTTTTCCACATTTGACAAATATATGAAATTCGAGTC 887
Qy      241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCys 260
Db      888 TTTCCAGATATAAGATATTCATTTGTTGGTTTCAATTCCTAGTAAAGTCAGCAGCATGCA 947
Qy      261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
Db      948 ATTGTTCTGTATATGTAATGTAACCAATGCAAGTCATGTTGAAATGCTATTTGGGCAAA 1007
Qy      281 AspGluLeuProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
Db      1008 GAAACTCT----- 1016
Qy      301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
Db      1017 --GATATGATATAATCC--GTGCACAGCAGCAATCAAAATGATATATCCC 1061
Qy      320 ----- 320
Db      1062 CAACCTATATGGCCAGTGGGCGCAGTGGTATGGAATGCAACAATAATGGCCGATATATG 1121
Qy      321 -----GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGlu 338
Db      1122 CCTATATGTTGGCAAGTTCCTGCATATGGAATGATGTCAGCAGGATGAAACGCAAGGA 1181
Qy      339 TyrAsnAspValValGlyLysGluProAsnTPrArgArgProSerAlaAsnGlyThrGly 358
Db      1182 TTTATCAAGACACAGCTCTTGCACCA---TGATGGGACCAAAAT-----TANGA 1229
Qy      359 IleLeuProSerProThrGlyPro-----GlyIleLeuProSerProAlaGlyGlyThr 376
Db      1230 GTGCACCG---CCTCAAGGGCAAAATGAGCAGCATGTGGCCCAATCAGCCTTCTGGGTAT 1286
Qy      377 Arg 377
Db      1287 CGA 1289

RESULT 10
US-07-843-949A-3
; Sequence 3, Application US/07843949A
; Patent No. 5340935
; GENERAL INFORMATION:
; APPLICANT: Anderson, Paul J.
; APPLICANT: Streuli, Michel
; APPLICANT: Schloesman, Stuart F.
; TITLE OF INVENTION: DNAS ENCODING PROTEINS ACTIVE IN
; LYMPHOCTYTE-MEDIATED CYTOTOXICITY
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:

```



```

ADDRESSER: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/218.978
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/843,949
FILING DATE: February 19, 1992
APPLICATION NUMBER: 07/460,678
FILING DATE: January 5, 1990
APPLICATION NUMBER: 07/726,607
FILING DATE: July 10, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/021003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1401
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-218-978-3

Alignment Scores:
Pred. No.: 2,14e-10 Length: 1401
Score: 242.50 Matches: 99
Percent Similarity: 39.29% Conservative: 68
Best Local Similarity: 23.29% Mismatches: 173
Query Match: 5.05% Indels: 85
Gaps: 21
US-09-920-705-3 (1-901) x US-08-218-978-3 (1-1401)
QY 4 SerMetLysProPheArgAlaAspSerGlyPheGlnSerAsnLeuTyrValGly 23
DB 31 TCGGTGCAACCCACCATGATGAGACGACGG--CAGCCCCGACCTTATACGTAGGT 87
QY 24 SerLeuThrProGluThrArgLysSerAspLeuThrGluLeuPheGlyArgTyrGly 42
DB 88 AACCTTCCAGAGATGTGACAGAGCTTATACCTTCACTTGTACGTACAGTTGACCC 147
QY 43 -----AspIleAspArgIleThrValTyrSerSerArgGly--PheAlaPheIle-- 58
DB 148 TGTAAAGCTGTAATGATTAACAGACATACAGACATGACCAATATGCTTGTGGA 207
QY 59 TTYTYAARGHIVaIglGluIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 78
DB 208 TTTTATGAACAC--AGAGATGAGCTGTGCTATGCTGCTATGATGGAATAAT 264
QY 79 AsnGlySerGlnIleLeuIleGluTyrAla--ArgProAlaIysProCysIysSer--- 96
DB 265 TTGGGAAAGAGAGGTAAAGTAACTGGGACCAACCAAGTAAAGCAAAAAGATATCT 324
QY 97 -----LeuTyrValGlyGlyIleGlyProAsnValSerIysAspAspLeu 111
DB 335 TCCAAATCACTTCAGATGCTTGTGTGGGATTTGAGTCCAGAAATTAACAAGAGATATC 384
QY 112 GluGluGluPheSerLysPheGlyLysIleGluAspPheArgPheLeuArgIuArgLys 131

```

```

DB 385 AATACGACATTTCCTTGTGTAATATGAGATGCCGGGTATGTAAGACATGGCA 444
QY 132 Thr-----AlaPheIleAspTyrTyrGluMetAspAlaLeuGln 145
DB 445 ACTGGAAATCCAAAGGCTATGTTTGTATCTTTTATTAACAACGATGACAAAAT 504
QY 146 Ala---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeu 164
DB 505 GCGATTTGCATATAGCGGTGACAGTGGTGGTGTCTCAATCCGAACCAATTTGGGCC 564
QY 165 -----ArgSerGlnAlaProLysLysGluIleTyrPalaGlySerTyrAspAsnArgAsn 182
DB 565 ACTCGTAACACCACTGCACCTTAAGTACACA-----GAAAACACACT 609
QY 183 GlyAsnMetAsnHisLysProGluTyrProHisSerTyrGluAspPheLysArgVal 202
DB 610 AAGCAGTTGAGA-----TTTGAAGATGTAGTAACCAAGTCA 645
QY 203 GlnProSerLys-----ValLeuTyrIleGlyPheProPheThrAlaThrGlnCysAsn 220
DB 646 AGTCCAAAATTTGACTGTGTACTGTGAGAGAAATTCGCTGGGTTAAC----- 696
QY 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
DB 697 ---GATGAGCTTATGAGACAGACATTCACCACTTTGGACAAATTATGAAATAAGATT 753
QY 241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgLysCys 260
DB 754 TTGCAGAAAGAGGGGTATTCATTTGTGATGATTCACCCATGAAAGTCCAGCCATGCC 813
QY 261 LysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsn 280
DB 814 ATGTGTTGGTGAACCGTACTACATTCAGAGACATGTGTTTAA---TGCTATTGGGCT 870
QY 281 AspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArg 300
DB 871 AAGATATCTCT-----GATATGACTTAAAACTTCCAAACGTTGACTATATGTCAA 921
QY 301 ThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIlePro 320
DB 922 TGG----- 924
QY 321 GlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGluTyrAsn 340
DB 925 GGCCAATGAGGCCAAGTGTATGA-----AACCACCAACGATATGA 966
QY 341 AspValValGlyLysGluProAsnTyrArgArgPro----- 352
DB 967 CAGTATATGGCAAT-----GGGTGCAAGTACCGCTTATGAGATACGGGCAACCA 1020
QY 353 ---SerAlaAsnGlyThrGlyIleLeuProSerProThrGlyPro-----GlyIle 368
DB 1021 TGGAAATCAACAGGATTTGGGTATGATCAATCACTTCTGCTGATGAGTGGATTT 1080
QY 369 LeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTyrGlu 388
DB 1081 GGTGCTCAGCTCCCAAGGCAAGCAAGTCCCTCCCTGTATACCTCTCTTAACAAGCC 1140
QY 389 GlyTyrAspProAla 393
DB 1141 GGATATGTATGCA 1155

```

RESULT 12
 US-09-514-247A-9
 ; Sequence 9, Application US/09514247A
 ; Patent No. 6365361
 ; GENERAL INFORMATION:
 ; APPLICANT: TANABE SEIYAKU CO. LTD.
 ; APPLICANT: TANIGUCHI, Tomoyasu
 ; APPLICANT: MIZUKAMI, Junko
 ; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND ANTAGONIST TO PPA
 ; FILE REFERENCE: TANIGUCHI=6

```

CURRENT APPLICATION NUMBER: US/09/514,247A
CURRENT FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: PCT/JP98/03734
PRIOR FILING DATE: 1998-08-24
PRIOR APPLICATION NUMBER: JP231084/1997
PRIOR FILING DATE: 1997-08-27
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 9
LENGTH: 8147
TYPE: DNA
ORGANISM: human
FEATURE:
NAME/KEY: CDS
LOCATION: (819)..(8147)
US-09-514-247A-9

Alignment Scores:
Pred. No.: 1,06e-08 Length: 8147
Score: 236.50 Matches: 203
Percent Similarity: 32.86% Conservative: 73
Best Local Similarity: 24.17% Mismatches: 297
Query Match: 4.93% Indels: 269
DB: 3 Gaps: 43

US-09-920-705-3 (1-901) x US-09-514-247A-9 (1-8147)

QY 251 PheArgSerAlaGluGluAlaArgGlnCysAlaGluGlnGlyArgLeuPheAsn 270
   |||||
DB 5189 TTTAGTAGTATGTGAAGAAATTGAGGTATGTGA-CAGGGCAATCTGGCGCTGCTC--- 5244
   |||||

QY 271 AsnProArgGlyLeuGlyMetTyrSerAsnAsp-----GluLeuPro 284
   |||||
DB 5245 -----CAAGTGAAGAGATGATTAATTCATCTTCATTCACCACTGATCAAAATAAC 5298
   |||||

QY 285 ProGluGlnAspAspThrSerPheTyrSerGlyMetLeuAspArg----- 298
   |||||
DB 5299 CCAAGCCAAACACACTGCGAG-----AGTGTACAAAAGATGCTGGACAGAGCGCTTG 5352
   |||||

QY 299 -----SerArgThrAspMetPheAsnAsnAspProSerCysValSerSer 313
   |||||
DB 5353 CAGAGCGGATCATCATGACTGACACAGATATTTTCAACAGCACTGAAGACAGGCTCA 5412
   |||||

QY 314 ProHisSerThrGlyLeuPro-----GlySerMet-----Arg 324
   |||||
DB 5413 CCAAGTCCAAAGAACTGCGCTTATTTTGAAGGTGATTTCTGCGCAATGTGTAGAAAGA 5472
   |||||

QY 325 ProLeuArgGlyThrAsnGluArgSerTyrAsnGlyValGluTyrAsnAspValAlaGly 344
   |||||
DB 5473 GCATTAAAGAACTTAGAACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5532
   |||||

QY 345 LysGluPro-----AsnTyrArg 350
   |||||
DB 5533 GTGAACCACTGAGGGGAGTCAGGGCGACAGCAAGATGCCAAGAAAGAACACAAAGA 5592
   |||||

QY 351 ArgPro-----SerAlaAsnGlyThrGlyLeuProSer---Pro 363
   |||||
DB 5593 AAACCAACAAAGAAACAAAGAGAGATCAGCGCGCCAAAGAAAGAACCCAGCATGCCCA 5652
   |||||

QY 364 ThrGlyProGlyLeuPro-----SerProAlaGlnGlyThrArgArg--- 378
   |||||
DB 5653 ACCTGTCCATGACCTGTCCAGAAAGCTGTATGCCACCATGAGAGAGACAAAGAGAGTCT 5712
   |||||

QY 379 -----ProMetArgSerAsn 383
   |||||
DB 5713 TCTTCGTGATCCACTGCAAGCTGAGCGCTGTATCAACACCTGCGCCCATGTGTGACC 5772
   |||||

QY 384 Pro-----AspSerTyrGluGlyTyrAspProAlaGlnLeuValArg 397
   |||||
DB 5773 CCAAGCCCGCTGCTCAGCTGATGACCTCATGATGGGCGCGACGCGCTTCTCTCA----- 5823
   |||||

QY 398 GluSerLysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAsp 417
   |||||

```

```

DB 5824 CCCTGCCAAGAGACAGCACTGGAGATTCTCTCTTGTGGCGGCTCCAAAGTGTCCAGCG 5883
QY 418 GluArgSerPheGlyArgGlySerValAlaAlaArgProIleArgGlyProPro----- 435
   |||||
DB 5884 TCTGCACTGCTGGAGAGCTGACACACCGAGGACAGACCGCTTGTCTACACTGCAACG 5943
QY 436 AspSerAspHisIleThrArgGlyMetIleAlaGlyGlyThrProValCysCys--- 454
   |||||
DB 5944 AGTCAAGACACACGTGGAGA-----CGCGTGGACCTGACTGATGTGGAGG 5991
QY 455 -----AlaArgCysValProMet----- 460
   |||||
DB 5992 ACTAGACCTCTGATCAATCTGCTATACAGCAAGACCATGCTCCATAGATGTGAAGT 6051
QY 461 ---GlyLeuGlyTyrIleGluThrLysLeuProGluValAlaAsnCysSerAlaArgThrAsp 479
   |||||
DB 6052 GGGGGCTGGGCTGTGATGACGAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6111
QY 480 LeuAsnMetLeu-AlaLysHisTyrAlaVal-----AlaIleGlyCysGluTyrLeuAlaPh 497
   |||||
DB 6112 AGGAGTCACCGCGCTGAGATCAGCGCTGATCCAGTCGCTGTGGTGCACGCTGCCA-- 6169
QY 497 ePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSe 517
   |||||
DB 6170 ---GTCCGCAACGC-----CAACTGCTGCTGCCATTCCT 6201
QY 517 rSerLysAspArgAlaGlyValAlaLysLeuAspArgGlyThrThrLeuPheLeuValPr 537
   |||||
DB 6202 GCCAGAAAGTGAAGCGGGTGTGCAGACACCAAGG----- 6238
QY 537 oProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyVa 557
   |||||
DB 6239 -----CTGCAACCGCAAGACCAACGGGGCTGCCGGGT 6273
QY 557 l-----ValLeuLysLeuProProAlaValProValThrAlaSerTyr 572
   |||||
DB 6274 GCAGAGAGCTCATCGGCTGCTGTCTACCGACCGCAAGACACTGCCAAGAAACAATGCC 6333
QY 572 rArg-----GlnGluSerGlnSerAsnProLeuHisTyrMetAspGlnAl 587
   |||||
DB 6334 CCGTGCCTTCTGCTCAATCAATCAACACAGCTCCGACGACGACGATCCAGCACCGCC 6393
QY 587 AlaArgSerProAlaAsnAla-----SerHisSerLeuTyrProProArgGluAs 604
   |||||
DB 6394 TGCAGAGCGCCAGCTCATGCGCGCGGATGGCCACATGAACACCCCAACGTCTTC 6453
QY 604 nTyrIleArgGlyAlaProGluHisIleuThrAlaAlaSerLysProSerValSerGlyPr 624
   |||||
DB 6454 A---GCAAGTCTGCTTCTCTACTTACGACCGCGGAGCCGCCACACAGACGCCCA 6510
QY 624 oLeuArgIleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeu 644
   |||||
DB 6511 G---CACACCCAGACGCGCGAGCCCTGCGCACAGCCCAACCTCACCCGTAG----- 6562
QY 644 uAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro----- 660
   |||||
DB 6563 ---CATGTCCACAGTGGCTTCCACGCGTGGCCGAGCTCGAATAAGCCACCACGCTGT 6618
QY 661 -----GluSerHisGlnProMetSerGlyProSerThrValAlaSerThrAlaIleGl 678
   |||||
DB 6619 CCACAGGAAAGCTTACACACAGTGCAGCGCGCCCGCCAC----- 6658
QY 678 nSerAsnGlyLeuTyrAsnGlyGluAlaProSer---GlnAlaTyrLysArgGlyProGln 697
   |||||
DB 6659 -----CCGCGCCAGCGCCCTCTGCAAGCGGTGAGAGCGGCT--CGGAG 6701
QY 698 ThrValHisAspAlaSerAsnGlnSer-----PheGlnGlnTyrGlyAsnGlnTyrThr 715
   |||||
DB 6702 ATTCAGCGTGAAGGCCAGAGAGACGACACTGTACCGGGTGAACATCAACAAACAGCATG 6761
QY 716 ProAlaGlyGln-----LeuProProProProSerArgTyrProProAlaSerAsnAsn 733
   |||||
DB 6762 CCCCCAGACGACGCGGATGGGGAGCCCGGGGAGGACAAATGGCCCTCGTGAAGCTGAAT 6821

```

```

QY 734 -----ProAsnTyrThrSerGlyMetValHisGlyAsnMet---GlnTyrGlnSer 749
Db 6222 GTGCCCCGACCCCAACGAGTGAAGCGGCGCCGTCAATGCCCAAGCTCCCGGAGTGG 6881
QY 750 GlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnTyrSer 769
Db 6882 CAGCAGCGCGCCCTCCCAAGCAGCAGCCATGCGCTTGGCCAGGCTGTGATATCC 6941
QY 770 MetTyrThrGlnGlySerSerAsnHisPro----- 779
Db 6942 ATGAGGCGCCAGCGCGCTGAGCTGCGCCCGATGCCAGCGGACCCAGAGAG 7001
QY 780 -----ValSerGlnProMet 784
Db 7002 ATCTCACCACCGCTCTGCAAGACCTCTGCGAGACCTGAAATGCGCCAGCTCCCTCAG 7061
QY 785 ValGlnGln----- 787
Db 7062 CAGCAACAGCAGGCTGTAACATTCTCAATCAAAACCGCAGCTAATGGCAGCTTCATC 7121
QY 788 -----TyrGlnProGlnAlaSerMet 794
Db 7122 AAACAGCGCAGACGCAAGTACGTGGCAATCAGCCCGCAGTGCAGCCCGCTGGCCTC 7181
QY 795 ProAsnGln---AsnTyrGlyProIleProSer---TyrGlnGlnAlaAsnPheHisGly 812
Db 7182 CAGTCCAGCGCCGCGCAGTCAACCCCGCAGCTGCGATGACAGCAGCCCGCAGCTG 7235
QY 813 ValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAsnMetGln--- 831
Db 7236 -----CAGAACCTGTAAT-----GCCATGCGAGCT 7259
QY 832 -----ProProAlaAspGlyAla-----AsnLeuGlnProGln 842
Db 7260 GCGGTCGCGCGCGCGCGCTGCTCCACAGCAGCAGCGCGAGCGCTGAAACCCCGAG 7319
QY 843 AsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThrAspGly 862
Db 7320 GGCAGGCGCTTGAACATCATG-----AACCCAGCAGCAGCAACCCCAAC 7361
QY 863 GluValAspLeuAsnGlnAlaGlyTyrGlnSerThrLeuGlnPheAlaAsnLeuLeu 882
Db 7362 ATGGCAGTATGTAATCCACAGTACCGAGAAATGTTACG-----AGCAGCTGCTGCAG 7415
QY 883 GlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 900
Db 7416 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 7469

RESULT 13
US-09-949-016-3886
/ Sequence 3886, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ FILE REFERENCE: C1001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 3886
/ LENGTH: 1968
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-3886

```

Alignment Scores:

Pred. No.: 6,22e-08
 Score: 215.50
 Percent Similarity: 36.95%
 Best Local Similarity: 23.33%
 Query Match: 4.49%
 DB: 4
 Gaps: 17

US-09-920-705-3 (1-901) x US-09-949-016-3886 (1-1968)

```

QY 11 AspaAspSerGlyPheGlnSerAsnLeuTyrValGlySerLeuThrProGlnThrThr 30
Db 291 GAGCAGAC-----AAGACCACTCATGCAACTACCTGCCCCAAGACATGACC 341
QY 31 GlnSerAspLeuThrGlnLeuPheGlyArgTyrGlyAspIle----- 44
Db 342 CAGATGAGTTCAAGATCTCTTGGCAGCATGGCGCATCGAGTCTCGCAAGTTGGTT 401
QY 45 ---AspaGlyIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyrArgHisVal 63
Db 402 CGGACCAAGATCACA---GGCAGAGCCTTGCTACGGATTGTGMACTATTCGACCCC 458
QY 64 GluGlnAlaValAlaAlaGlyGlnAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIle 83
Db 459 ATATGATGACAGCAAGCCATCAACACCTTCAACGGCTCAATTAACAGACAAACATC 518
QY 84 IysIleGluTyrAlaArgProAlaLysProCys-----LysSerLeuTyrValGly 100
Db 519 AAGGTGCTTATGCGACAGCCAGTTCACATCCAGTTCAGGATGCTAATCTGTACGTGAC 578
QY 101 GlyIleGlyProAsnValSerLysAspAspLeuGlnGlnGlnPheSerLysPheGlyLys 120
Db 579 GGGCTCCCAAGACCATAGCCAGAAAGATGAGAGCCTTCTCCAGTACGCGCGC 638
QY 121 IleGluAspPheArgPheLeuArgLysThrAlaPhe-----IleAspTyr 137
Db 639 ATCATCAGCTCCCGCATCTGTGTGACAGGTCAACAGTCTCTCGGGGTGTGGATTC 698
QY 138 TyrGlnMetAspAspAlaLeuGlnAla-----LysSerMetAsnGly---Lys 152
Db 699 ATCCGCTTTGACAGAGGATGAGCGCCGAGAGGCTTCAAAAGATGGAATGGCAGAAAG 758
QY 153 ProMetGlyLysSerPhe---LeuArgValAspPheLeuAsnSerGlnAlaProLysLys 171
Db 759 CCGCTGGGCGCAGCTGAGCCCATCAAGTTCGGAACAACCAAGTCAGAGACG 818
QY 172 GluGln-----TyrAlaGlySerTyrAspAsnArgAsnGlyAsnMet 185
Db 819 GGGCAGGCGCTGCTACCCACCTCTACCAATCTCCGCGGCTTACGAGGCCCTCTA 878
QY 186 AsnHisLysProGln----- 190
Db 879 CACCATCAGACCGCGCTTCCGGCTGAGACATTTGCTCAATGAGCTTACGCGTCAAG 938
QY 191 ---TyrProHisSerTyrGluAspPheLysGly-----AspVal 202
Db 939 AGCTTTCGCGCATGCCATGATGATGATGAGCGGCTGGCGGCGCTGTGCGGG 998
QY 203 GlnProSerLysValLeuTyrPheIleGlyPheProThrAlaThrGlnCysAsnAspGlu 222
Db 999 GCGCGCGGCGCGCGCGCTGGTGCATCTTCGTGTAACAACGTTCACCGAGGCGAGAG 1058
QY 223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGlnIleGlnArgValLys----- 239
Db 1059 AGCGTGTGTGGCAGCTGTTCGGGCTTTGGGCACTCAACACGTCAAGTTCATCCGT 1118
QY 240 ---SerTyrProSerArgAsnPheAlaLeuValGlnPheArgSerAlaGln 256
Db 1119 GATTTCACACCAACAGTGAAGGTTTGGCTGTGTGACATGACCACTATGACGAG 1178
QY 257 AlaArgGlnCysLysGlnGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIle 276

```

Db 1179 GCGCCATGCGCATCGCAGCCTGAACGCTATCGCTGGGAGCGGTGTCAGATC 1238
Qy 277 MetTyrSerAsnAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMet 236
Db 1239 TCCTTC----- 1244
Qy 297 LysArgSerArgThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSer 316
Db 1245 AATACACGAAACAGCAAGGCGGTGAGCCACCCCGCTGCC--TCCACCCC----- 1297
Qy 317 ThrGlyIleProGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGly 336
Db 1298 -----CTCCCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1351
Qy 337 AlaGluTyrAsnAspValValGlyValGluProAsnTrpArgArgProSerAlaAsnGly 356
Db 1352 GCCCAAGAGAGAGAG-----CAAGGACAGCCCGCAGAGAGAGAGAGAGAGAGAG 1390
Qy 357 ThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThr 376
Db 1391 -----GAGGCGCCACAGCTCCTGCGGAAGCCA---CAGGTGAG 1426
Qy 377 ArgArgProMetArgSerAsnProAspSerTrpGluGly 389
Db 1427 CACTCTGGGTGGAGAGGTCTGAGAGGAAATTGGGGGGGT 1465

RESULT 14
US-09-410-551B-26
Sequence 26, Application US/09410551B
Patent No. 6503737
GENERAL INFORMATION:
APPLICANT: KOSAN BIOSCIENCES, Inc.
APPLICANT: REEVES, CHRISTOPHER
APPLICANT: CHU, DANIEL
APPLICANT: KHOSLA, CHAITAN
APPLICANT: SAMTI, DANIEL
APPLICANT: WU, KAI
TITLE OF INVENTION: POLYMERIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
TITLE OF INVENTION: CONSTRUCTS THEREFOR
FILE REFERENCE: 30062-20026.00
CURRENT APPLICATION NUMBER: US/09/410,551B
PRIOR APPLICATION NUMBER: US 60/139,650
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/123,810
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 60/102,748
PRIOR FILING DATE: 1998-10-02
NUMBER OF SEQ. ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26
LENGTH: 4674
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
OTHER INFORMATION: PKS synthase fragment
FEATURE:
NAME/KEY: CDS
LOCATION: (3)..(4673)
US-09-410-551B-26

Alignment Scores:
Pred. No.: 1,01e-06 Length: 4674
Score: 208.00 Matches: 209
Percent Similarity: 31.27% Conservative: 109
Best Local Similarity: 20.55% Mismatches: 400
Query Match: 4.33% Indels: 300
DB: 4 Gaps: 44

US-09-920-705-3 (1-901) x US-09-410-551B-26 (1-4674)
Qy 10 AlaAspAspSerGlyPheGlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThr 29

Db 1743 GCGGACAGCGCGGTATTCAGAGGGGACCGGAGTGGGTCCCGTGGTATTCGGCCAGG 1802
Qy 30 ThrGluSerAspLeuThrGluLeuPheGlyArgTyr-----Gly 42
Db 1803 ACCAGTCCGCTTTGACTGATGACGAGGCGCGGTTCGCGTATTCGCGCGGTGCC 1862
Qy 43 AspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyrArgHis 62
Db 1863 GGGGTGATATGCGGCGCTGGCATGACCTGGCGATGACACCGTGGTGTTCGAGCAC 1922
Qy 63 -----ValGluGluAlaValAlaAlaValGluAla 72
Db 1923 CGTCCGCTGCTGTGGAGATGACACCGTCACCGGACCGGCTGTGTCTGACCTCGGGG 1982
Qy 73 Leu-----GlnGlyAlaAsnLeuAsnGlySerGlnIleValIleGlyTyr 87
Db 1983 GTGTTCGCTTCCTCCGAGACAGGCGGTGCGCGTGTGCGATGGGTGAGAGAACTGGCGGC 2042
Qy 88 AlaArgPro-----AlaArgProCysArgSerLeuTrpValGlyGlyIleGlyProAsn 105
Db 2043 GCGTTCCTCCGCTCTTCGCGGATTCATCAGCAGGTGTG-----GACCTG 2087
Qy 106 ValSerTyrAspAspLeuGlu----- 112
Db 2088 CTCGATGTCCTCCGATCTGAGAGGTGAAAGAGACCGGTTAGCCCGCCCTGTTCGCA 2147
Qy 113 GluGluPheSerTyrPheGlyIleValIleGluAspHe----- 124
Db 2148 ATGACAGTGGCTCTGTTCCGCGCTGCTGTAATCGTGGGTGTGTACACCGACCGGTGATC 2207
Qy 125 ---ArgPheLeuArgGluArgIleGlyThrAlaPheIleAsp---TyrTyrGluMetAsp 142
Db 2208 GGCATTCGTGGGTGATGCTTCGCGCTGTGATGTCCGGGCTGTGTGTGGAGAT 2267
Qy 143 Ala-----LeuGlnAlaValSerMetAsnGlyValProMetGlyGlySer 157
Db 2268 GCTGTGACTTTGGTGTGCGCGCGGCTCTGTGATGACAGCTTCGCCCGGTGGGG 2327
Qy 158 PheLeuArgValAspPheLeuArgSerGlnAlaProValGluGlnIleTrpAlaGlySer 177
Db 2328 ATGTCGCTGTC----- 2339
Qy 178 TyrAspAsnArgAsnGlyAsnMetAsnHisValProGlnTyrProHisSerTyrGluAsp 197
Db 2340 -----CCGCTTCGAGAGATGAG 2357
Qy 198 PheValGlyAspVal---GlnProSerTyrValLeuTrpIleGlyPheProProThrAla 216
Db 2358 GCGCGGCGCTGTGGGTGAGGCTGTGAGATCGCGCGGTCAACGCGCGCTGTGCGTG 2417
Qy 217 ThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGlu 236
Db 2418 GTTCTCTCCGATGATGACCGCGCTGTCGAGCGCGGAGGCGGTGGGAAGTGAGG 2477
Qy 237 ArgValIleSerTyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGlu 256
Db 2478 CGGCTGCGACAGCAGCCAGCTTCATTCGCGCGGTATGAA---CCCATGCTGAGAG 2534
Qy 257 AlaArgGluCysLeuGlyLeuGluIleGluGlnIleGluValArgLeuPheAsnAspProGlyIleVal 276
Db 2535 TTCGCGGCGGTGCGCGAGAGCTG-----ACCTACCGAGCGCGAGGTCTCAGT 2585
Qy 277 MetTyrSerAsnAspGluLeuProProGlu-----GlnAspAspThr 290
Db 2586 GCGGTGTGATGATGATGACACCGCTGATGCTGCTGCGGACGAGTCCGAGACAGGTC 2645
Qy 291 SerPhe-TyrSerGlyMetValArgSerArgThrAspMetPheAsnAsn----- 306
Db 2646 CCGTTCGAGACAGGTGCGCTGTATCGAGACGCGGTGCTGTGAGCTGGGTGCGGAC 2705
Qy 306 ----- 306

Dp	2706	CGGTACTGTCGGCCGCTGGTTCGACGGTGTGCGGAGTGTCTGCACCGGAGCACCAAGAAATCCAG	2766
Qy	307	-----AspProSerCysValSer-----	312
Dp	2766	GCCGGATCGCGCCCTGTGGCCACCTGTATGTCAACGGCGTCACTGCAGTCGGCCCG	2825
Qy	313	-----SerProHisSerThr	317
Dp	2826	CTTCCTGGCGGATGCTTCGCGAAACAGGGTCTGGAACCTTCGACATTAAGCTTTCACGAC	2885
Qy	317	rgIyIleProGlySerMet-----ArgProLeuArgGlyThrHisenGluAr	332
Dp	2886	CAGCGCTACTGTCTGAGTGGCTCCCGCGGACAGGCGGACATCGGGGACACCCGTCCTC	2945
Qy	332	gSerTyraenGlyIaGluTyraAsnAspValaIyI-----	344
Dp	2946	GGCACCGGAGTCCCGCTGCGCGGGTCCGGGCGGGGCTTTCACAGGATCCCGTGGCCGC	3005
Qy	345	-----LysGluProAsnThrArg-----	350
Dp	3006	GGTGGCGAGCCGCGCGTGTATATCGCCGAATCGGGCTCTCGCCCGCGCGACGCCAC	3065
Qy	351	-ArgProSerAlaenGlyThrGlyIleleuProSerProThrGlyProGlyIleuPr	370
Dp	3066	TGGCGCACAGTGTGCAACAGTGCAGCTCACTCCGTCGCCCGCGGATCCG-----	3114
Qy	370	oSerProIaenGlyI-----ThrArgArgProMetArgSerAsnProAspSerTyrGluI	389
Dp	3115	-----CCCGCGAGGGGCGCACCGCGGACCTGGGTGATGAACCGCGCGCGGACGGGCG	3170
Qy	389	Y-----TyrAspProIaIaGlnLeuValArgIuSerIySaGthTrArgIaGAs	405
Dp	3171	CGCCGCTTCACCGTCCACACCGCGTGGCGACCGCCCGGTGAGCGCTGCACGCGGAGG	3230
Qy	405	pgIySerValaIaSpGly-----PheThrProMetGlyValaIaSpGly	418
Dp	3231	GTTCTCGCGCCCGCGCGGTGCGCCAGCGCGCAAGCCGTGCACACCGCTGGCCCCCGCG	3290
Qy	418	uArgSerPhe-----GlyArgGlySerValaIaIaIaArgProIaAr	432
Dp	3291	GGCGCGGTGCGCGCGAGCGGGCTGCCGGGCGGTGGCGACCGCGGACCAAGTCTTCGTC	3350
Qy	432	gGlyProProAspSerAspHisIleTyrArgGlyMetIleIaIySgIyIThrProVa	452
Dp	3351	GAAAGCCGAAGTCGACA-----GCCCTGACGGCTTCGTGGCACACCCCGAC	3395
Qy	452	lCyeCySaIaArgCysValaProMetGlyIySgIyIleGluThrIySleuProGluVala	472
Dp	3396	CTGCTCCGACGGGTCTTCTCCGCG-----TCGGCGAC	3428
Qy	472	IaSnCySerAlaArgThrAspLeuAsnMetLeuAlaIyHisTyraIaValaIaIeGly	492
Dp	3429	GAGAGCCGCCAGCGGACCGGATGG-CGGACCTCGCGGTGAC-----	3470
Qy	492	YCySgIuIleValPhePheValaProAspArgGluGluAspPheIaIaSerTyThrGluPh	512
Dp	3471	-----GCGTCGACGCGCACCGT	3487
Qy	512	eLeuArgTyIleuSerSerIyIyAspArgIaGlyValaIaIyLeuAsp-----AspGly	530
Dp	3488	GCTGGCGCCCTGCTCACCGCGCGGACAGTGTGTGTGTGGAGCTCGCGCGCTTTCGACGG	3547
Qy	530	YThrThrIeu---PheLeuValaProProSerAspPheLeuThrAspValaIeGluIaIth	549
Dp	3548	TGCCGGAATCCGCGGTCTCACCGGAGATCGGTACGCTGGCGGAGGTTCGGTTCGAGG	3607
Qy	549	rArgGlnIuArgLeuTyrgIyValaIal-----LeuIySleuProPro-ProA	565
Dp	3608	CGGATCCGACAGTCCGACCGGTCTGCTTCGCGCTTGAAGTGTTCGCGGTGGCGGAGGCCA	3667
Qy	565	IaValaProValaThrIaIaSerTyraIyArgGlnIuSerGlnIyAsnProLeuHisTyIyMetA	585
Dp	3668	CTACAGCGGTCCGACAGAGCTGGCCGAGAGGCTACACCTTATATCCGCGCACACCCCGA	3727

Oy 585 spglnlaIarGAsperProAlasnaIsaSerHisSerLeuTy-ProAlargluAuant 605
 Db 3728 CGACCCCGCAGCACCCCAACCCCAACAACAACACCACACGCCACCCACACAACAACAC 3787
 Oy 605 yrIleArGIyAlAProglunHIsleuthralala-----SerlyeProSeValSerg 623
 Db 3788 ACGCGTCTCAACGGCTTCACAAACACCACTCATACACCAACCAACCACTCATGTCTCA 3847
 Oy 623 luProleuArGIleProshasnmlalalProglinalaglyValSerleuthrProglul 643
 Db 3848 CACCA-----CACCGACCCCCAGGCGCGCGGTCAACCGGCTCAACCCGCACCGCA 3901
 Oy 643 euleuAlaThrleuAlaserlleuProAlathnSergInProAlalalProgluseR 663
 Db 3902 AAAGGAACACCCCGCGCCA-----TCCACTCATGMAAACCAACCAACCCCAACACCC 3955
 Oy 663 IsglnPmet-----SerglyProSerThr-ValValSerThralahlgInserN 680
 Db 3956 ACTCCCTCCATCCCAACTCACACCTCTCAACCAACCCCACTTACGCTCACCAACA--- 4012
 Oy 681 glyleuTyrannglylualProSerglnlalatrpylsarAglyProglInThValHis 700
 Db 4013 -----CACCTCCACACCCCACTCAACCCCAATCAACCAACCAACCAACACACAC 4066
 Oy 701 AsplaseAenInserPhehgInglyngInlyrtyrValngInlyrThrProalaglyIndeu 720
 Db 4067 AAC-----CACCCCAACACCCCAAC 4087
 Oy 721 ProProProProSerArgTyProProAlasernaSnProAntTyTrhsSerglyMet 740
 Db 4088 CCTCAACCCCAACCAACCCCACTCTCAT----- 4114
 Oy 741 ValHsglysnMetGInTyrgInserdInserValasnMetProglInleuSerPro--- 759
 Db 4115 ---CACGGCGGCTCGGCACTCGCGGGATCTCGCCGCACTCAACACCCCA 4171
 Oy 760 ---LeuProasnMetProHisasnAntyrSerMetTyTrhgInglySerSerAshis 778
 Db 4172 CACCTACTCTCTCTCGGCAACACCAACCCCAACACCCG-----CAC 4219
 Oy 779 ProValserglnPmetValGInglyrInProglualasernMetProashnlnasn 798
 Db 4220 CCA-----CATCCCCG--CGACCTCACGACCCCAACCAATCACCAAGGCTTCAC 4270
 Oy 799 TyrclyPro-IleproserTyrgIngnlnalasnPhHgIyValthr----- 815
 Db 4271 -----COAcTYACCAACCCCTCACCGGATTTCCACACCGCGCCACCTCGACGA 4324
 Oy 816 -AsnglnalaglnasnleuanProSerglnPhehglnalalawetGlnProProAlas 835
 Db 4325 CGCACCTCAACCAACTCAACCCCAACACTCACACCAACCTCCAAACCAAGCGA 4384
 Oy 835 pLyvala-----AsnleugInProglInAsnglnaleu 846
 Db 4385 CGCGGCTTGACCTCCACCAACCAACCAACCAACCCCTC 4427

RESULT 15
 US-09-940-316B-26
 Sequence 26, Application US/09940316B
 Patent No. 6759536
 GENERAL INFORMATION:
 APPLICANT: KOSAN BIOSCIENCES, Inc.
 APPLICANT: REEVES, CHRISTOPHER
 APPLICANT: CHU, DANIEL
 APPLICANT: KHOSLA, CHAITAN
 APPLICANT: SANTI, DANIEL
 APPLICANT: WU, KAI
 TITLE OF INVENTION: POLYPEPTIDES ENCODING THE FKBP GENE OF THE PK-520 POLYKETIDE SYNTHASE
 FILE REFERENCE: 30062-20026.11
 CURRENT APPLICATION NUMBER: US/09/940.316B
 CURRENT FILING DATE: 2001-08-27

PRIOR APPLICATION NUMBER: 09/410,551
PRIOR FILING DATE: 1999-10-01
PRIOR APPLICATION NUMBER: US 60/139,650
PRIOR FILING DATE: 1999-06-17
PRIOR APPLICATION NUMBER: US 60/123,810
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 60/102,748
PRIOR FILING DATE: 1998-10-02
NUMBER OF SEQ ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26
LENGTH: 4674
TYPE: DNA
ORGANISM: Artificial Sequence
OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
OTHER INFORMATION: PKS synthase fragment
FEATURE:
NAME/KEY: CDS
LOCATION: (3)..(4673)
US-09-940-316B-26

Alignment Scores:
Pred. No.: 1,01e-06 Length: 4674
Score: 208.00 Matches: 209
Percent Similarity: 31.27% Conservative: 109
Best Local Similarity: 20.55% Mismatches: 400
Query Match: 4.33% Indels: 300
Gaps: 44
US-09-920-705-3 (1-901) x US-09-940-316B-26 (1-4674)
QY 10 AlaAspAspSerGlyPheGlnSerAsnLeuTrpValGlySerLeuThrProGluThr 29
DB 1743 GCGACACACCGCGTATCGAGCGGCGCACCGAGTGGCGCGTGGATTCGCCAGG 1802
QY 30 ThrGluSerAspLeuThrGluLeuPheGlyArgTyr-----Gly 42
DB 1803 ACCGAGTCGGCTTGACTGACGACGAGGCGCGTGGCGTATCTGGCGCGTCCGCC 1862
QY 43 AspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyrArgHis 62
DB 1863 GGGGTGATATGGCGGCTGTGGCATCGACCTGGCCATGACAGCTGGCGTTCGAGAC 1922
QY 63 -----ValGluGluAlaValAlaAlaValGluAla 72
DB 1923 CGTGGCGTCTGTGGAGATGACACCGTCACCGGCGACCGCTGTGTGACCTCGGCG 1982
QY 73 Leu-----GlnGlyAlaAsnLeuAsnGlySerGlnIleValIleGluTyr 87
DB 1983 GTTTCGTCTTCCCGGACAGGGGTGCGACCGTGTGGCATGGGTGAGAACTGGCGCC 2042
QY 88 AlaArgPro-----AlaTyrProCysLeuSerLeuTrpValGlyIleGlyProAsn 105
DB 2043 GCGTTCCTCCCTCTTCCGCGCGATCCATCAGCAGTGTG-----GACCTG 2087
QY 106 ValSerIysAspAspLeuGlu-----GACCTG 2087
DB 2088 CTCGATGTGCGCGATCGAGGTGAAGACAGCGGTTACCGCCAGCGCGCTGTTCGA 2147
QY 113 GluGluPheSerIysPheGlyLysIleGluAspPhe-----124
DB 2148 ATGACAGTGGCTCTGTTCGGGCTGTGGAATCGGGGTGTACGACCGACGCGGTGATC 2207
QY 125 ---ArgPheLeuArgGluArgLysThrAlaPheIleAsp---TyrTyrGluMetAspAsp 142
DB 2208 GCGCATTCGGTGGGTGAGCTTGGCGCTGCTGATATGTCCGGGTGTGTGTTGAGAGAT 2267
QY 143 Ala-----LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySer 157
DB 2268 GCGTCGACATTGTGTGTGGCGGCGCTGTGTGATGACAGGCTCGCGGCGGAGGTG 2327
QY 158 PheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGluGlnTrpAlaGlySer 177

DB 2328 ATGTCGCTGTC-----2339
QY 178 TyrAspAsnArgAsnGlnIysAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAsp 197
DB 2340 -----CCGCTCTCGAGAGATGAG 2357
QY 198 PheLysGlyAspVal---GlnProSerLysValLeuTrpIleGlyPheProProThrAla 216
DB 2358 GCGCGGCGCGTGTGGGTGAGAGGTGTGAGATGCGCGGCTCAACGCGCTGTGCTGCTG 2417
QY 217 ThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGlu 236
DB 2418 GTTCTCTCCGGTATATAGGCGCGCTGTGACAGCGCGGAGGCGTGGGAGATGAGAC 2477
QY 237 ArgValLysSerTyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGlu 256
DB 2478 GCGCTGGCGACACGACCGCGTTCATTCGCGCGATGGA---CCATGTCTGAGAGAG 2534
QY 257 AlaArgGlnCysLysGluGluLysLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIle 276
DB 2535 TTCCGGCGCGTGCACAGAGCGCTG-----ACCTACCGACGCGCAGGTCTTCATG 2585
QY 277 MetTyrSerAsnAspGluLeuProProGlu-----GlnAspAspThr 290
DB 2586 GCGCTGTGTATAGGTGACACCGCTGTGATGCTGAGTGGCGGAGTCCGGACAGCGTC 2645
QY 291 SerPhe-TyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsn-----306
DB 2646 CGGTTGCGGAGAGAGTGGCGCTGTGACAGAGACCGGTGTGTGAGCTGGGTGCGGAC 2705
QY 306 -----306
DB 2706 CGGTCACTGACCGCGCTGTGACAGGTGTGCGGATGCTCACGCGACCAAGAAATCCAG 2765
QY 307 ---AspProSerCysValSer-----312
DB 2766 GCGCGATGGCGCGCTGTGACCTGTATATGACAGCGGTCAAGCTCACTGAGTCCCGCG 2825
QY 313 -----SerProHisSerThr 317
DB 2826 CTCCTGGCGAGATCTCCGGAACACGAGGTGTGACCTTCCGACATATGCTTCCAGAC 2885
QY 317 GlnIleProGlySerMet-----ArgProLeuArgGlyThrAsnGluLys 332
DB 2886 CAGCGTACTGCTGATGATGCTGCTCCCGCGCACCGGCGACCTCGGCGCACCGCTCTC 2945
QY 332 GSerTyrAsnGlyAlaGluTyrAsnAspValValGly-----344
DB 2946 GGCACCGAGTGCCTGCGCGGCTGCGCGGCGGAGGTTCACGGGTCCCGTCCCGCC 3005
QY 345 -----LysGluProAsnTrpArg-----350
DB 3006 GGTGGGAGACCGCGGCTGTTCATCGCGAATCGGCGCTGCGCGCGCGACCGCACCGAC 3065
QY 351 -ArgProSerAlaAsnGlyThrIleLysLeuProSerProMetGlyProGlyIleLeuP 370
DB 3066 TGGCCACGATCGAAGAGCTGACGTCACTGCTGCTGCTGCGCGGATTCG-----3114
QY 370 oSerProAlaGlnGly---ThrArgArgProMetArgSerArgAsnProAspSerTrpGlu 389
DB 3115 ---CCCGGCGAGGCGCACCGGCGACCTGTGCTGATGAAACCCCGCGCACGCGGCG 3170
QY 389 Y-----TyrAspProAlaGlnLeuValArgLysSerLysArgThrArgArgHis 405
DB 3171 CGCGGCTTACAGGTTCACACCGCGCTGCGGAGACCGCGGTGACGCGCGAGGAG 3230
QY 405 pGlySerValAspGly-----PheThrProMetGlyValAspGlu 418
DB 3231 GTTCTCTCGCGCGCGCGCTGCGCGACCGGAGCGGTGACACCGCTGCGCGCGCGCG 3290
QY 418 wArgSerPhe-----GlyArgGlySerValAlaAlaArgProIleArg 432

Db 3291 GGCGGGTGCCTGGCGAGCGGGCTGCGGGGCGTGGCGAGCGCGGAGCAGAGCTTCTGTC 3350
 QY 432 GGGYProProAspSerAspHisIleTrrArgIleMetIleAlaIleGlyIleProVal 452
 Db 3351 GAAGCCGAAGTCGACA-----GCCCTGACGGCTTCGTGGCAGCACCCCGAC 3395
 QY 452 ICGYsAlaIleArgCysValProMetGlyIleGluThrIleLeuProGluValVal 472
 Db 3396 CTGCTGACGGCGGTCTTCTCCGG-----TCGGCGAC 3428
 QY 472 IAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaIleHisIleTrrAlaIleIleG 492
 Db 3429 GGGAGCGCGCAGCGGACCGGATGG-CGCGACCTCGGGTGAC----- 3470
 QY 492 YCysGluIleValPhePheValProAspArgGluGluAspPheAlaSerTrrThrGluP 512
 Db 3471 -----GCCCTCGAGCCGACCGT 3487
 QY 512 eLeuArgTrrLeuSerSerIleAspArgAlaGlyValAlaIleLeuAsp-----AspG 530
 Db 3488 GCTGGCGGCTGCTGCTACCGCGCGAGCAGTGGTGTGGAGACTGCGCGCTTCAGACG 3547
 QY 530 YrrThrLeu---PheLeuValProProSerAspPheLeuThrAspValIleGluIleVal 549
 Db 3548 TGGCGGAATGCTCGGTCTCACCGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3607
 QY 549 rArgGluIleArgLeuTrrGlyValVal-----LeuIleLeuProPro- ProA 565
 Db 3608 CGGATCCGAGAGTGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3667
 QY 565 IValIleProValThrAlaSerTrrArgGluIleSerIleAsnProLeuHisIleTrrMet 585
 Db 3668 CTACGACGGTGGCGAGAGTGTGCGAGGGCTACACCTTCATCACCGCACACACCCGA 3727
 QY 585 spGlnIleArgAspSerProAlaAsnAlaSerHisSerLeuTrrProProArgGluAsn 605
 Db 3728 CGACCCGAGAGCCGACCCCAACCCCAACACACACCGCACACCAACACAC 3787
 QY 605 YrrIleArgIleAlaProGluHisLeuThrAlaAla-----SerIleProSerValSerG 623
 Db 3788 ACCGCTCTCACCGCTCTCCACACACACCTTCATCACACACACACACCTTCATGCTCA 3847
 QY 623 IuProLeuArgIleProAsnAsnAlaIleProGluIleGlyValSerLeuThrProGlu 643
 Db 3848 CACCA-----CCACCGACCGCCGAGCGCGCTCACCGGCTCACCGCGACCGCAC 3901
 QY 643 euleuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaIleProGluSerH 663
 Db 3902 AAACGAACACCCCGCGCA-----TCACCTCATCGAAACCCACACCCCGACACCC 3955
 QY 663 IeGlnProMet-----SerGlyProSerThr-ValIleSerThrAlaHisGlnSerAsn 680
 Db 3956 ACTCCCTTCACCCACTCACCACTTCACCAACCCCACTACGCTTCACCAACAA----- 4012
 QY 681 GlyLeuTrrAsnGlyIleAlaProSerGlnAlaTrrIleArgIleProGluThrValHis 700
 Db 4013 -----CACCTTCACACCCCGACCTCACCCCGATCACCCACACCAACACACAC 4066
 QY 701 AspAlaSerAsnIleSerPheGlnGlnTrrGlyAsnGlnTrrThrProAlaGlyIleLeu 720
 Db 4067 AAC-----CACCCCGACACCCCGAC 4087
 QY 721 ProProProProSerArgTrrProProAlaSerAsnAsnProAsnTrrThrSerGlyMet 740
 Db 4088 CCTCAACCCCAACACCGCATCTCAT----- 4114
 QY 741 ValHisGlyAsnMetGlnTrrIleSerGlnSerValAsnMetProGluIleLeuSerPro--- 759
 Db 4115 ---CACCGCGGCTCGGACCTCGCGCATCTCGCGCGCACCTCAACACACCCCA 4171
 QY 760 ---LeuProAsnMetProHisAsnAsnTrrSerMetTrrThrGlnGlySerSerAsnHis 778
 Db 4172 CACCTACCTCTCTCTCCGACACACACACCCCGACACACACCGG-----CAC 4219

QY 779 ProValSerGlnProMetValGlnGlnTrrGlnProGluAlaSerMetProAsnGlnAsn 798
 Db 4220 CCA-----CATCCCTG---GACCTCACGACCCCAACCCCAAAATCACCCCAAGCCCTCAC 4270
 QY 799 TrrGlyPro-IleProSerTrrGlnGlnAlaAsnPheHisGlyValThrThr----- 815
 Db 4271 -----CCATTCACCAACACCCCTCACCGGCACTTTCACACCGCGCCGACCTCGACGA 4324
 QY 816 -AsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaIleMetGlnProProAlaAs 835
 Db 4325 CGCACCTTCACCAACCTTCACCCCGCAACCTTCACACACCTTCACCAACCAAGCCGA 4384
 QY 835 PysAla-----AsnLeuGluProGlnAsnGlnAlaLeu 846
 Db 4385 CGCGCTGTGGACCTTCACACACCAACCAACCAACCCCTC 4427

Search completed: April 8, 2005, 13:52:02
 Job time : 453 sec8

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 12:01:35 ; Search time 1224 Seconds

(without alignments)
4460.836 Million cell updates/sec

Title: US-09-920-705-3

Perfect score: 4801
Sequence: 1 MALSMKPRFRDSDGFSNNL.....LQIQKQKQSSGSTRAGCP 901

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5615251 seqs, 3030001701 residues

Total number of hits satisfying chosen parameters: 11230502

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h
-Q=/cgn2_1/USPTO.spool/h/US09920705/runat_07042005_101816_20810/app_query.fasta_1.1095
-DB=published Applications NA -QFMT=fastbp -SUFFIX=rmh -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cst -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09920705@cgn_1_1_622@runat_07042005_101816_20810
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA: *
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCF_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCFUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4801	100.0	2706	10	US-09-920-705-2
2	4392	91.5	4593	10	US-09-920-705-1
3	1436	29.9	801	10	US-09-920-705-4
4	1286.5	26.8	3463	18	US-10-425-115-123950
5	1281.5	26.7	3084	18	US-10-437-963-36321
6	802	16.7	3315	18	US-10-434-963-79215
7	669	13.9	1101	17	US-10-424-599-59704
8	651.5	13.6	3715	17	US-09-920-705-6
9	499	10.4	801	17	US-10-424-599-65572
10	466	9.7	349	10	US-09-920-705-5
11	463	9.6	1630	17	US-10-425-114-23888
12	447	9.3	1631	17	US-10-424-599-40095
13	406	8.5	2159	17	US-10-425-114-31541
14	382	8.0	815	18	US-10-767-701-2544
15	359	7.5	606	18	US-10-767-701-30057
16	336.5	7.0	6836	18	US-10-475-917-1
17	333.5	6.9	1146	17	US-10-424-599-18149
18	309	6.4	729	17	US-10-424-599-47401
19	302	6.3	483728	18	US-10-699-156-2
20	297	6.2	1621	18	US-10-425-115-118399
21	276.5	5.8	3383	18	US-10-425-917-11
22	273	5.7	3312	18	US-10-475-917-7
23	271.5	5.7	10656	17	US-10-475-917-9
24	270	5.6	1241	17	US-10-425-114-35702
25	268	5.6	3228	17	US-10-302-117-752
26	265	5.5	10995	17	US-10-263-929-41
27	265	5.5	12227	15	US-10-177-293-422
28	265	5.5	12227	18	US-10-684-422-243
29	265	5.5	11295	10	US-09-814-353-15153
30	261	5.4	683	17	US-09-770-445-972
31	258.5	5.4	438	9	US-09-834-975-1036
32	254	5.3	1871	9	US-10-172-118-410
33	246.5	5.1	2228	17	US-10-342-887-410
34	246	5.1	2228	17	US-10-723-860-6794
35	246	5.1	2482	18	US-10-425-115-79600
36	241.5	5.0	443	18	US-10-109-886-9
37	236.5	4.9	8147	13	US-09-938-842A-1801
38	236.5	4.8	1803	9	US-09-938-842A-1801
39	229.5	4.8	1803	11	US-10-263-929-40
40	225	4.7	16883	17	US-10-425-115-118394
41	223.5	4.7	2201	18	US-10-425-115-31949
42	218.5	4.6	2340	18	US-10-723-860-8366
43	216	4.5	2774	18	US-10-425-115-25600
44	212	4.4	3433	18	US-10-425-115-25600
45	212	4.4	3433	18	US-10-425-115-25600

ALIGNMENTS

RESULT 1
US-09-920-705-2
Sequence 2, Application US/09920705
Publication No. US20030079252A1
GENERAL INFORMATION:
APPLICANT: Amasino, Richard M.
APPLICANT: Schomburg, Fritz M.
APPLICANT: Michaels, Scott D.
APPLICANT: Patton, David
TITLE OF INVENTION: Floral Induction Gene
FILE REFERENCE: 960296. 97214
CURRENT APPLICATION NUMBER: US/09/920, 705
CURRENT FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 2706
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:


```

QY      621 ValSerGluProLeuArgGluProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThr 640
      1861 GTTAGCGAGCCCTCCAAATTAATCAATGAGCGCTCAAGCTGGGTTAGTTAACT 1920
QY      641 ProGlnLeuLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaPro 660
      1921 CCGAGGCTTTTAGCACTCTGGCATCTATCTCCCTGCAACTTCAACTGCTGCCCT 1980
QY      661 GluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsn 680
      1981 GAGAGTCACCAACCTATGTCAGGACCTTCAACAGTTGTTCCACAGACATCATGCTCAAT 2040
QY      681 GlyLeuThrAsnGlyGluAlaProSerGlnAlaThrLysArgGlyProGlnThrValHis 700
      2041 GGAAGTCAATGAGAGAACCCGCTCAAGCTTGAAAAGAGTCCAAACAGATTTCAT 2100
QY      701 AspAlaSerAsnGlnSerPheGlnGlnIleValAsnGlnIleThrProAlaGlyGlnLeu 720
      2101 GATGCGTCAATCATGATTCCTCAACAAATACGAAATACATCACTCCAGCTGGGCAACTA 2160
QY      721 ProProProSerArgIleProProAlaSerAsnAsnProAsnIleThrSerGlyMet 740
      2161 CCTCTCTCTCTCGCTTACCTCCAGCTTCAAAACCCCAACTACATGATGAGATG 2220
QY      741 ValHisGlyAsnMetGlnIleValIleSerGlnSerValAsnMetProGlnLeuSerProLeu 760
      2221 GTCCATGAGCAACATGCAATACGAGAGCAATCTGTTAACTGCTCCAGCTGCTCCGTA 2280
QY      761 ProAsnMetProHisAsnAsnIleValIleSerMetIleThrGlnIleSerSerAsnIleProVal 780
      2281 CCAATATGCTCTCAATATTAATTTCCATGATACATGAGGTTCTGCAATCATCTCTT 2340
QY      781 SerGlnProMetValGlnGlnIleValIleProGlnAlaSerMetProAsnGlnAsnIleVal 800
      2341 TCTCAGCCCAATGCTCCAGCAATACCAACGAGCGCTTCAACCAAACTATGCT 2400
QY      801 ProIleProSerIleValIleValIleAsnPheHisGlyValIleThrAsnGlnAlaGlnAsn 820
      2401 CCAATTCCAATGATACGAGCAAGCTTAATTTCAATGCGCTAACCAATTCAGGACAGAAC 2460
QY      821 LeuAsnProSerGlnPheGlnAlaAlaMetGlnProProAlaAspLysAlaAsnLeuGlu 840
      2461 TTAAACCTTCCCAATTTCAAGCTGCCATGCCAACACAGACAGATTAAGGAAATTTTAG 2520
QY      841 ProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGlyAspGlyGlnIleThrThr 860
      2521 CCACAAAACCAAGCACTACGATTCAGAGCTATGATCTCTGGGAGTGGTCAAGGTTACACA 2580
QY      861 AspGlyGlyValAspLysAsnGlnIleValIleSerThrLeuGlnPheAlaAlaAsnLeu 880
      2581 GATGGGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2640
QY      881 LeuLeuGlnIleGlnIleValIleValIleValIleValIleValIleValIleValIle 900
      2641 CTTCTCCAGATACGACGAAACAGACAGACAGATTTTCAAGTACTCGCGCTGACAGGG 2700
QY      901 Pro 901
      2701 CCT 2703

```

RESULT 2

```

US-09-920-705-1
; Sequence 1, Application US/09920705
; Publication No. US20030079252A1
; GENERAL INFORMATION:
; APPLICANT: Amasino, Richard M.
; APPLICANT: Schomburg, Fritz M.
; APPLICANT: Michaels, Scott D.
; TITLE OF INVENTION: Floral Induction Gene
; FILE REFERENCE: 960296, 97214
; CURRENT APPLICATION NUMBER: US/09/920, 705

```

```

; CURRENT FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4593
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-920-705-1

Alignment Scores:
Pred. No.: 0 Length: 4593
Score: 4392.00 Matches: 899
Percent Similarity: 58.72% Conservative: 0
Best Local Similarity: 58.72% Mismatches: 2
Query Match: 91.48% Indels: 631
DB: 10 Gaps: 4

US-09-920-705-3 (1-901) x US-09-920-705-1 (1-4593)
QY      1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnAsnLeu 20
      1 ATGGCGTTATGATGAAAGCCATTGACAGCGGATTCGGGTTTCCAGTCAACAAATCTT 60
Db      1 ATGGCGTTATGATGAAAGCCATTGACAGCGGATTCGGGTTTCCAGTCAACAAATCTT 60
QY      21 TPValGlySerLeuThrProGluThrGlnIleValIleSerAspLeuThrGluLeuPheGlyArg 40
      21 TGGGTCGTAGCTTAACGCCGAGACGACAGAGTCAATCTGACCGAGTTGTTGAGAGA 120
Db      61 TGGGTCGTAGCTTAACGCCGAGACGACAGAGTCAATCTGACCGAGTTGTTGAGAGA 120
QY      41 TyrGlyAspIleAspArgIleThrValIleYserSerArgIlePheAlaPheIleIleIle 60
      41 TACGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
Db      121 TACGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
QY      61 ArgHisValGlyGluAlaValAlaAlaValIleValIleValIleValIleValIleVal 80
      61 AGACATGTGAGAGAGACAGTCCAGCCAAAGAGCTCTTCAAGAGCAAAATTTGAATGA 240
Db      181 AGACATGTGAGAGAGACAGTCCAGCCAAAGAGCTCTTCAAGAGCAAAATTTGAATGA 240
QY      81 SerGlnIleLysIleGlyIleValIleValIleValIleValIleValIleValIleVal 90
      81 AGTCAAAATTAATGAAATACGACAGACCGGTTTGTCTTATCTATCTGTTGTTTC 300
Db      241 AGTCAAAATTAATGAAATACGACAGACCGGTTTGTCTTATCTATCTGTTGTTTC 300
QY      90 ----- 90
Db      301 TCTAATCTGATGCTTTTGTCAAGATTAATCTTTTTCGAATCATAGTCCAGGT 360
QY      90 ----- 90
Db      361 TCACAAACTTGATGATGCTTTGTTAGTCAAAAATTTCTGTGAACTGTTTTTTT 420
QY      90 ----- 90
Db      421 TTCCTCATGTACAAATCAAACTGCAAACTGTTTTTTTCTATTATACGTGTTAGCTT 480
QY      90 ----- 90
Db      481 AAGGCAAACTGATCCGATGCAAAAGCTTTTCTCAATTAATCTTTGTTATATGCAACT 540
QY      90 ----- 90
Db      541 CCGCAAAAGCAAAACACAGAGAGCTTCAAAATTTGATGTTAAAGCATATATACTC 600
QY      90 ----- 90
Db      601 TTAGCAATGAGCTTCGAAAAGAAATACATCAACACATGTTTACTCTGTTTATGCGAA 660
QY      90 ----- 90
Db      661 GAAGTTTAATCCGATTTGCTTTATCTGCAAACTTGTATCGCGTTTACAGCTGATTT 720
QY      90 ----- 90
Db      721 TCAGTTGCTTCAAGTGTAAATCTCAACAGCTTGAATGAATGATATGCGTACTCA 780
QY      90 ----- 90

```

Db 781 CTCGAATTGCAATGGTGAATAATTGGAAAGAGCAAACTATCATCGGCCAAACAGA 840
 QY 90 ----- 90
 Db 841 TAAGAAATTTGGAGTTTAAAGTTCAAGTTCTCGCAAAAATCAACCCTGGAGAAAT 900
 QY 90 ----- 90
 Db 901 TTGTATAGCGGAGTAGATATCAATTAATGATCTGCGAAACAGAAATTTCTGGCAA 960
 QY 90 ----- 90
 Db 961 CCGTATGCTTAATCTTATGATATCGCTGACATATCCATATAGTGTGAGGAGTTT 1020
 QY 90 ----- 90
 Db 1021 GGGCTTGTGGCTGATGCTTCAAGAGACAAATTAAGTGTGTAAAGCGGCTAAC 1080
 QY 90 ----- 90
 Db 1081 TCTACCAATGAGAAAGCTTTATTCGAAGAACATGTTGTCTCAATTCACATCCCTA 1140
 QY 90 ----- 90
 Db 1141 CGTACAATCTGGGCTTTCCCATTTAGTTCCTTAGGAAGTTGACTTCTTCAACACCATTC 1200
 QY 90 ----- 90
 Db 1201 TGGGATTTGTACAGTTTGCATGAGCAAGACAAATATTCATTTGACAGCAGACTCAATTT 1260
 QY 90 ----- 90
 Db 1261 CAATGCTTGGCGGCTTTGAGTACTTAAATGAGTGTTCAGGATTTGTTATTGGGAC 1320
 QY 90 ----- 90
 Db 1321 TCAGCAATGAGTAAGAACAGAAATGTTTGCAGAAAGTAATATGTCCTTCGCGCTTGG 1380
 QY 90 ----- 90
 Db 1381 GGGACTACATACTTCTCTTAAGACAAATCTGATGTTTAGGACAGAGAACTATTACAA 1440
 QY 90 ----- 90
 Db 1441 CTGATTTAAAGAAAGAGAGAGATGTTTCTCTTCGCGCAATTTTATTGTTAGAA 1500
 QY 90 ----- 90
 Db 1501 AGCAATTGATGAATGGTGTCTGATGCTAGTTGAATTAATTAAGTTAGTTGTGTGTT 1560
 QY 90 ----- 90
 Db 1561 AGTTCTTGATGTTGATGCTTTATTTCTTGCAACCTATCGGGATGATGCCCTTCT 1620
 QY 91 ----- AAlaLysProCysLysSerLeuTyrValGlyGlyIleGly 103
 Db 1621 ATCGACCTTTTCTTGTGGCAGGCAAAACCTGTAAAGCTATGAGGTGGGTGGAAATCGGC 1680
 QY 104 ProAsnValSerLysAspAspLeuGluGluIlePheSerLysPheGlyLysIleGluAsp 123
 Db 1681 CCTAAATGCTCCAAAGATGACCTGGAGAGAGTTCAAGAAATTTGGGAAATCGAGGAT 1740
 QY 124 PheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
 Db 1741 TTTAGGTTTCTCGAAGACGCAAGACAGCTTTCATTTATTTAAGATGATGATGCT 1800
 QY 144 LeuGluAlaLysSerMetAsnGlyLysProMetGlyLysSerPheLeuArgValAspPhe 163
 Db 1801 TTACAGAGCTTAAGACATGATGAGAAAGCTATGGGTGTAGCTTTTGTGCTGTGATTTT 1860
 QY 164 LeuArgSerGluAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLys 171
 Db 1861 CTCGGCTCAAGAGCCCAAAAAGTAAGCACTTTGTGGCAATTGATTTTACTTTTGA 1920

QY 171 ----- 171
 Db 1921 AAAGCTCCAGTAAACATTTGTTAGTTTCAATAATTTGGCTCAACATGATAGGCTGAG 1980
 QY 171 ----- 171
 Db 1981 CTCTGTCTTGTGCCCTAGAGACATATTACTGCTCTATTTGATTTGATAGAGCT 2040
 QY 172 ----- GluGlnTyrAlaGly 176
 Db 2041 CAATCTTAAAGTCTGAATAATCAAGTTACCTTGTGTATCTTCAGAGAAATAGGCTGAC 2100
 QY 177 SerTyrAspAsnArgAsnGlyAsnMetAsnHisLysProGln----- 190
 Db 2101 TCTTAAGATACAGAAATGGCAATATGATATTAACCGCA-GGTTAGCTTGAATGTTG 2159
 QY 190 ----- 190
 Db 2160 AAAGTATGCTCTGTCTAGTATGATATGATATAGTTACTGTGTTGACGTTTGTATTA 2219
 QY 191 ----- TyrProHisSerTyrGluAspPheLysGlyAspValGlnProSerLysVal 207
 Db 2220 TTCTTACAGTATCTCTCACTCATATGATGAAAGCTTTAAAGAGATGTCCAGCCAAAGTAAGGCT 2279
 QY 208 LeuTyrPileGlyPheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsn 227
 Db 2280 CTGTGATTTGGGTTCTCTCTCTACTGCTACACAAATGCAATGATAGCAAAATTTCTGCACAT 2339
 QY 228 AlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrProSerArgAsnPheAla 247
 Db 2340 GCGATGATATCTTTGATGAGATCGAGGGTAAAGATTAATCCATCAAGAAATTTTGCA 2399
 QY 248 LeuValGluPheArgSerAlaGluLysAlaArgGlnCysLysGlyLysGluGlnIleValArg 267
 Db 2400 CTGTGAGATTTAGAGGCGGAGAGAGAGCTCGCAATGCAAGAGGCTTACAGGGGAGG 2459
 QY 268 LeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAspGluLeuProProGluGln 287
 Db 2460 TTATTCAATTAATCTCGAATCAAAATTAATGATCTCAACAGATGAGTTGCTCTCGAGCAA 2519
 QY 288 AspAspThrSerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAspAsp 307
 Db 2520 GACGATACTAGTTTATCTCTGATGATAAGAGGTCAGAGACAGATATGTTCAATTAATGAT 2579
 QY 308 ProSerCysValSerSerProHisSerThrGlyLysProGlySerMetArgProLeuArg 327
 Db 2580 CTTCAATTGTATCTTCTCTCATTTCTAAGTAAATCTGGGTCTATGAGGCCCCCTCAGA 2639
 QY 328 GlyThrAsnGluArgSerTyrAsnGlyValAlaLysTyrAsnAspValValGlyLysGluPro 347
 Db 2640 GGTAAATATGACGCTCATATATATGTCAGAAATACATATGATGCTGTTGGTAAGAGCCA 2699
 QY 348 AsnTyrArgArgProSerAlaAsnGlyThrGlyLysLeuProSerProThrGlyProGly 367
 Db 2700 AACTGAGAGAGGCAATCTGCAATGAACTGAATATCTCCCATCTCCAAACGAGACTGGA 2759
 QY 368 IleLeuProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTyr 387
 Db 2760 ATCCCTCCATCTCTGCAAGAGTACAGGCGCCCTTAAGAGTAAACCCCGATTTCTGG 2819
 QY 388 GluGlyTyrAspProAlaGlnLeuValArgGluSerLysArgThrArgArgAspGlySer 407
 Db 2820 GAAGATATATGATCTCTGCTGACTGTGTGTCAGAGAAATGAACAAACAGAAAGATGATCA 2879
 QY 408 ValAspGlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySerValAla 427
 Db 2880 GTGACGCTTTTACTCCAAATGAGGTGTGATGACAGAGCTCATTTGTGACAGGTTCAAGTGTCT 2939
 QY 428 AlaArgProIleArgGlyProProAspSerAspHisIleTyrPargGlyMetIleAlaLys 447
 Db 2940 GCTAGACCTATCCGAGGCCCCCTGATTTCTGATCAATATGAGAGAGAAATGATTTGCCAG 2999


```
Db 739 ACACAATGCAATGATGACAAATTCGACAAATGCAATGATCTTCTTGGTGAATCGAG 680
QY 237 ArgVallySerSerTyProSerArgAsnPhaAlaLeuValGluPheArgSerAlaGluGlu 256
Db 679 AGGGTAAAGATTACCATCAAGAAATTTTGCACTTGAGATTGAGACCGGGAGAA 620
QY 257 AlaArgGlnCysIysGluGlyLeuGlnGlyArgLeuPheAsnProArgIleVal 276
Db 619 GCTCGCAATGCAAGAGAGCCCTACAGGAGGATTTCTCAATATCTTAGAATCAAAATT 560
QY 277 MetTySerAsnAspGluLeuProProGluGlnAspAspThySerPheTySerGlyMet 296
Db 559 ATGTACTCAACAGCATGAGTTGCTCCCTGACAGCAAGCATACTAGTTTCTCTGGTAG 500
QY 297 LysArgSerArgThyPheMetPheAsnAspProSerCysValSerProHisSer 316
Db 499 AAACGGCAGAGACAGATATGTTCAATATGATCTTCATTTGTATCTCTCTCATTTCT 440
QY 317 ThrGlyIleProGlySerMetArgProLeuArgGlyThyAsnGluArgSerTyAsnGly 336
Db 439 ACTGGAATTCCTGGGTATAGAGCCCTCAGAGGTAACAATAGCGTTCAATTAATGGT 380
QY 337 AlaGluTyAsnAspValValGlyLeuGluProAsnTyrArgArgProSerAlaAsnGly 356
Db 379 GCAGAAATACAAATGACGTTGTTGGTAAAGAGCCAAACTGAGAGAGCCATCTGCAATGGA 320
QY 357 ThrGlyIleLeuProSerProThyArgGlyProGlyIleLeuProSerProAlaGlnGly 376
Db 319 ACTGGAATTCCTCCATCTCCAAAGAGACCTGGAATCTCCATCTCTGACAAAGGTAG 260
QY 377 ArgArgProMetArgSerAsnProAspSerTyrGluGlyTyAsnProAlaGlnLeuVal 396
Db 259 AGCGCGCTATAGAGTCAACCCCGATTCCTGGGAAGATATGATCTCTCTCATTTGGT 200
QY 397 ArgGluSerIleArgThyArgArgArgGlySerValAspGlyPheThyProMetGlyVal 416
Db 199 AGGAAGATMAAGAACCCAGAAAGATGATGATCAGTACGACGTTTACTCCAAATGGGTGTC 140
QY 417 AspGluArgSerPheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAsp 436
Db 139 GATGAGAGTCAATTTGGTCAAGGTTCAAGTTGCTGCTAGACCTATCCGTGACCCCTGAT 80
QY 437 SerAspHisIleTyrArgGlyMetIleAlaIysGlyGlyThyProValCysCysAlaArg 456
Db 79 TCTGATCAATATGAGAGGAATGATGCCAAGGTGGAACCTCCGTGTTGTGTCTGT 20
QY 457 CysValProMetGlyLys 462
Db 19 TGTGTACTATGGGAAAG 2

RESULT 4
US-10-425-115-123950
; Sequence 123950, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 123950
; LENGTH: 3463
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_44523C.1
US-10-425-115-123950
```

```
Alignment Scores:
Pred. No.: 1,666-106
Score: 1286.50
Percent Similarity: 49.80%
Best Local Similarity: 35.63%
Query Match: 25,80%
DB: 18 Gaps: 40

US-09-920-705-3 (1-901) x US-10-425-115-123950 (1-3463)
QY 16 GlnSerAsnAsnLeuTyrValGlySerLeuThrProGluThrThyGluSerAspLeuThr 35
Db 227 GAAAGCAACAGCTCTGGGTGGGCAACTGCCCTCGACAGTGAACGAGGCGGATTTGTG 286
QY 36 GluLeuPheGlyArgTyGlyAspIleAsp--ArgIleThrValTySerSerArgGly 54
Db 287 GCGCTGTTCCGGCCCGCAGCGCCGCTCATTTGCCCTTGCCGACGAGCCGATCTCGCAGC 346
QY 55 PheAlaPheIleTyrTyArgHisValGluGlnAlaValAlaIysGluAlaLeuGln 74
Db 347 TAGCGGTTGCTCTCTCCGTTCCCGACCGAGGCTCGGAGCCGTCGAGGCGACCCGG 406
QY 75 GlyAlaAsnLeuAsnGlySerGlnIleIysIleGluTyrAlaArgProAlaLysProCys 94
Db 407 GGTGAAAAGTCAAGGGGGCGCCGATGGCAACGAGTTCCGGCGCGCGCTAGAGCTGTT 466
QY 95 LysSerLeuTyrValGlyGlyIleGlyProAsnValSerLysAspAspLeuGluGlu 114
Db 467 AGAAATCTGTGGTGTGGGATTAAGCCATTCCTCAAGAGAGAGCTTGAGGGGAG 526
QY 115 PheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPhe 134
Db 527 TTTCTAAAGTTCCGAAAAGTTGAGAGTGTGCTTCTCCCAAGATCAACTTCACGATAC 586
QY 135 IleAspTyTyTyGluMetAspAspAlaLeuGlnAla--LysSerMetAsnGlyLysPro 153
Db 587 ATAGATTTTGAAGAGCTTGAAGAGCAATTTCTGTCACAGATCTTGAATGAAAAATG 646
QY 154 MetGlyGlySerPheLeuArgValAspPheLeuArgSerGlnAlaProLysGluGln 173
Db 647 CTAGGTGGCAAGAAATTTGTGTGATTTCCAGAGGTCACAAAGGAGAGACCGAATGTGCA 706
QY 174 TrpAlaGlySerTyTyAspAsnArgAsnGlyAsnMetAsnHisLysProGlnTyrProHis 193
Db 707 GAAGTTACAGCTTCAATGATGA-----GTACCAAGG 739
QY 194 SerTyGluAspPheLysGly----- 200
Db 740 CCAGTGGGTGATTAAGAGAGGAGTGGCCCTCGAAGATTCGTGAGTGGGATGCGA 799
QY 201 AspValGlnProSerLysValLeuTyrIleGlyPheProThrAlaThyGlnCysAsn 220
Db 800 GAGGACAACTTACCAATGTTCTCTGGGTGGT---CCAGGTTCTTATGAGTCATT 856
QY 221 AspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSer 240
Db 857 GATGAGAGGAGCACTTATGACATGCAATTCGATTCGTCCTTACGAAATAAAGTT 916
QY 241 TyrProSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGlnAlaArgGlnCys 260
Db 917 TTCCAGACAAAGCAGTATGCTTTGTGAGTTTCAATATGTTGGGAAAGCTTGAATGCT 976
QY 261 LysGluGluLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTySerAsn 280
Db 977 AAGATGAATCTTAGATGTCATCTTTTCAATATGCCAGATTCAGATTCCTTTTCAAC 1036
QY 281 AspGluLeuProProGluGln--AspAspThySerPheTySerGlyMetLysArgSer 299
Db 1037 AGTGGGCTTGACAAACAAATTTGACACCAACCAATCGGTGTGATTTCTCAATGTTCA 1096
QY 300 ArgThyAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThyIle 319
```

Db 1097 -----GATATCTATTCAGTAGTGTGCTGTAGAGCCCTGTATTTGGCAGTGGAACTTTG 1150
 Qy 320 ProGlySerMetArgProLeuArgGlyThrAsnGlnArgSerTyrAsnGlnAlaGluTyr 339
 Db 1151 CAAGGG---TATGATTCACCAAGAGAGAGAGAAACAAG---TAT 1189
 Qy 340 AsnAspValAlaGlyIleGluProAsnTyrArgArgProSerAlaAsnGlyThrGlyIle 359
 Db 1190 TAGGATTTATGCGGGT----- 1204
 Qy 360 LeuProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgPro 379
 Db 1205 -----ATGCCCTACTCCAGGTGGATCTTTCCAAACCT----- 1237
 Qy 380 MetArgSerAsnProAspSerTyrGluGlyTyrAspProAlaGlnLeuValArgGluSer 399
 Db 1238 -----GAACCATTTGATCCA-----AAGAAGACA 1261
 Qy 400 LysArgThrArg-----ArgAspGlySerValAsp 409
 Db 1262 AAAAGATGAGCTGTGCTGTGCTGACCTTACATGTAAAGGAGGACAGCAGCAGT 1321
 Qy 410 -----GlyPheThrProMetGlyValAspGluArgSerPheGlyArgGlySer 425
 Db 1322 CTTTATTCCTGCTGATACCGTCATCAGGGCAGTTCTGTGACCGCTGAGGAGAGCTCAAGT 1381
 Qy 426 ValAlaAlaArgProIleArgGlyProProAspSer-----AspHisIleTyr 441
 Db 1382 CCGTATTATTCGA---GTCCGGGGCAGCAGTGCATCGTACATCATCATCATGAGCATGTGG 1438
 Qy 442 ArgGlyMetIleAlaLysGlyGlyIleThrProValCysAspAlaArgCysValProMetGly 461
 Db 1439 CGTGGCAGATTTCGCAAAAGGTGATCCCTGTTGTGCTGCTGCTGTTGCTTATTAACA 1498
 Qy 462 LysGlyIleGluThrLysLeuProGluValAlaAsnCysSerAlaArgThrAspLeuAsn 481
 Db 1499 AAGGGCAGTGCATACCTTTACCGGATGTATGAATTTGCTCAGCTCGAGCTGATTTGAT 1558
 Qy 482 MetLeuAlaLysHisTyrAlaValAlaIleGlyCysGluIleValPhePheValProAsp 501
 Db 1559 ATGCTGCGCAAGACATACCGGATGCGCATGATTTGACATTTGCTTTTTCGCAAGAT 1618
 Qy 502 ArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSerLysAspArg 521
 Db 1619 AGCGAAGATGACTTGTGTTCTTATCTGAGTCTTGCGTTACCTGGGCTCAAAAGTCGG 1678
 Qy 522 AlaGlyValAlaLysLeuAspAspGlyThrThrLeuPheLeuValProProSerAspPhe 541
 Db 1679 GCAGGGGTGTGTAAGTGTATGACAGGGACTACATTTATTTTGGTCCACCATCAGATTTTC 1738
 Qy 542 LeuThrAspValLeuGlnValThrArgGlnGluArgLeuTyrGlyValValLeuLysLeu 561
 Db 1739 CTAAACAATGTGTAAAGTGTATGATGTCCAGACGCGCTTATGTGTGTATTAACATTT 1798
 Qy 562 ProProProAlaValProValThrAlaSerTyrArgGlnGlu-----SerGlnSer 578
 Db 1799 -----CCACAATCTCTGCTGCTGCTGACCTAGGCGCACAGTTAACTGGAGCCGAACA 1852
 Qy 579 AsnProLeuHisTyrMetAspGlnAlaArgAspSerPro---AlaAsnAlaSerHisSer 597
 Db 1853 CAACCT-----TACTATGATGAAGAGGGAATCTGCTACTTACACAAAGAAAGTATAGT 1906
 Qy 598 LeuTyrProProArgGluAsnTyrTyrIleArgGlyAlaProGluHisLeuThrAlaAlaSer 617
 Db 1907 ATCATTTCTCTATATGACAGT-----GGCCACCTTGATGTGATAT 1948
 Qy 618 LysProSerValSerGlu-----ProLeuArg----- 626
 Db 1949 CGGACATCTTTTGATGATAGATTAATGATCATCGTGGGGCATATTTCTTGAGAGGCTCGG 2008
 Qy 627 -----IleProAsnAsnAlaAlaPro 633
 Db 2009 GTGATGAAGGCCAAGCAGTTCAACAGACTCTTGAGGTTTCTCTATATCAAGCAGCT 2068

Qy 634 -----GlnAlaGlyValSerLeuThrProGluLeuLeuAlaThrLeuAla-SerIleIle 651
 Db 2069 GGGCTGCAAGTACATCTTCAAGCTCAAGCTTATATGCTTACATTTGGCAAAACCTTCT 2128
 Qy 651 uproAlaThrSerGlnProAlaAlaProGluSerHisGlnProMetSerGlyProSerThr 671
 Db 2129 GCTTAGTGTG-----CAATGTGCACCGCTGTAGTAAAGTCAAGTAA 2170
 Qy 671 rValValSerThrAlaHisGlnSerAsnGlyLeuTyrAsnGlyGluAlaProSerGlnAl 691
 Db 2171 TGCCATAGACAGAAACATCCGAG-----ATACGAGATCCATCCATCTTTCAAAAGT 2221
 Qy 691 aTrrLysArgGlyProGlnThrValHisAspAlaSerAsnGlnSerPheGlnGlyTyrG 711
 Db 2222 ATGG-----AATCCGAAACCAAGTATCTGCTTCAAAATTAATCTTTGGCAAAATGGC 2275
 Qy 711 yAsnGlnTyrThrProAlaGlyGlnLeuProProProProSerArgTyrProProAlaSe 731
 Db 2276 TATATGTTCAGATCCAGACAGCAGATTT-----ATGGCCAAAGCTTCAGCTCG 2323
 Qy 731 rAsnAsnProAsnTyrThrSerGlyMetValHisGlyAsnMet---GlnTyrGlnSerG 750
 Db 2324 TCATTTGACAAACTAT-----GGAAACATGTGAGTGTACAGAGGCG 2365
 Qy 750 nSerVal-----AsnMetProGlnLeuSerPr 759
 Db 2366 CTCATATACAACTACTCTTACACCCCTGAAGTACTTTGAACCTTGCCACCAACCTTC 2425
 Qy 759 oLeuProAsnMetProHisAsnAsnTyrSerMetTyrThrGlnGlySerSerAsnHisPr 779
 Db 2426 ACTTCCAAACAATACCAACAGCTTGTCTATCAATCAAGTCAAGTGGCAGCAAGTTGGC 2485
 Qy 779 oValSerGlnProMetValGlnGlnTyrGlnProGluAla---SerMetProAsnGlnAs 798
 Db 2486 C---TCACAGACAAACAGCAACTGATATCAGACAGAGCATATATGTATGTCGCCCAAAACA 2542
 Qy 798 nTyrGlyProIleProSerTyrGlnGlnAlaAsnPheHisGlyValThrThrAsnGln-- 817
 Db 2543 CTATGCTCATTTAGTTCAGTTAGCCATTTACCTTCAATCAGCAACATCAACACCC 2602
 Qy 818 -----AlaGlnAsnLeuAsnProSerGlnPheGlnAlaIleMetGlnProProAl 834
 Db 2603 CACCTTACCATCCCAACAGTAAACCTTGA-----CTTCCAC 2641
 Qy 834 aAspLys-----AlaAsnLeu---GluProGlnAsn-----GlnAlaLeuArg 847
 Db 2642 AATATATCAGATTGGGAATTTGGCCAGCCACAGCATTTCTATGCCACTGATTTGATAG 2701
 Qy 847 gLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThrAspGly----- 862
 Db 2702 ACCAAGTCAGATTTCTTCTCAGGGGCAACAAATCGGTGCTCGTGGTCGACACA 2761
 Qy 863 -----GluValaLysLysAsnGlnArgTyrGlnSerThrLeuGlnPheAlaAlaAsnLe 880
 Db 2762 AGCTCCCAAGAGGATTAAGAGAAAGATACAGAGCAGACATCTCCAGTTAGCTCAAAACCT 2821
 Qy 880 uLeuLeuGlnIleGlnGlnLeuGlnGlnGln-----SerSe 893
 Db 2822 ACTGCTTCAGTTACAGCAGCGTGATCTGAATCAATCTTGAGACTTATATTATGCTTC 2881
 Qy 893 rGlyThrProAlaGlyGlnGly 900
 Db 2882 AGGTACCTCGTCTGCTGTGT 2903

RESULT 5

US-10-437-963-36321
 ; Sequence 36321, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei
 APPLICANT: Mu, Wei
 APPLICANT: Boukharov, Andrey A.
 APPLICANT: Barbazuk, Brad
 APPLICANT: Li, Ping
 TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53221)B
 CURRENT APPLICATION NUMBER: US/10/437,963
 NUMBER OF SEQ ID NOS: 204966
 SEQ ID NO 36321
 LENGTH: 3084
 TYPE: DNA
 ORGANISM: Oryza sativa
 FEATURE:
 OTHER INFORMATION: Clone ID: PAT_MRT4530_40159C.1
 US-10-437-963-36321

Alignment Scores:

Pred. No.:	4,06e-106	Length:	3084
Score:	1281.50	Matches:	339
Percent Similarity:	51.47%	Conservative:	150
Best Local Similarity:	35.68%	Mismatches:	306
Query Match:	26.69%	Indels:	155
DB:	18	Gaps:	38

US-09-920-705-3 (1-901) x US-10-437-963-36321 (1-3084)

QY 16 GlnSerAsnAsnLeuTrpValGlySerLeuThrProGluThrGlnSerAsnLeuThr 35
 DB 178 GAAACCAACACGCTGTGGGGGCAACCTGCCCGCAGCGCGGAGAGACGCGTGG 237
 QY 36 GluLeuPheGlyArgTrpGlyAspIleAspArgIleThrValTyrSerSer---ArgGly 54
 DB 238 GGCCTCTTCCCGCCGACGCGCGCTGACCTGCGTATGACCCGCGCGGCGCGGAGC 297
 QY 55 PheAlaPheIleTyrTrpArgHisValGluGluAlaValAlaAlaGlyGluAlaGln 74
 DB 298 TAGCGCTTCGTGCTGTTCCGCTCCGTCGCGGAGCGCGCGCGCTCGACCGCTCCAG 357
 QY 75 GlyAlaAsnLeuAsnGlySerGlnIleValIleGluTyrAlaArgProAlaLysProCys 94
 DB 358 GGGTCCAGAGTCAGAGGACGCGCTGCGCTCGAGTTCGCGCGCGCGCTGAGCTGTT 417
 QY 95 LysSerLeuTrpValGlyGlyIleGluProAsnValSerLysAspAsnLeuGluGln 114
 DB 418 AGGAATTATGGGTTGGTGGCATTTAGCTCTCAATTTCAAGAGAGAGCTGGAGAGAG 477
 QY 115 PheSerLysPheGlyLysIleGluAspPheArgPheLeuArgGluArgLysThrAlaPhe 134
 DB 478 TTTAAGAAATTGGAAAGGTGATGATGATTCCTCGTGAACAGACTTCCCGGAC 537
 QY 135 IleAspTyrTyrGluMetAspAsnAlaLeuGlnAla---LysSerMetAsnGlyLysPro 153
 DB 538 ATTGATTATGACAGGTGAGAGATGCAATTTGGGACATAGACCTTGAATGGAAGAGT 597
 QY 154 MetGlyGlySerPheLeuArgValAspPheLeuArgSerGlnAlaProLysLysGluGln 173
 DB 598 TTAGGTGAGCAAGAACTGTGTGTTGATTTCCAAAGTCCAGGGGC-----AGAAGGAA 651
 QY 174 Trp-----AlaGlySerTyrAspAsnArgAsnGlyAsnMetAsnIleLysProGlnTyr 191
 DB 652 TGGTTGAGACAGGAGCTTCAATGTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 696
 QY 192 ProHisSerTyrGluAspPheLysArgValAspGlnProSerLysValLeuTrpIleGly 211
 DB 697 -----GGTTATGAGAGTACAAATGAGATCCAAATCCCAATGTTCTTTGGGTAGT 750
 QY 212 PheProProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeu 231
 DB 751 TTCCTTAATACAGCTTAA-----ATCAATAGAGAGGACATTGGCAAGCAAGTGGCTGTA 804

QY 232 PheGlyGluIleGluArgValLysSerTyrProSerArgAsnPheAlaLeuValGluPhe 251
 DB 805 CATGGTGTGTACAAATACAAATAGTTTCCCAACAGCAATATGCTTTGTGAGATT 864
 QY 252 ArgSerAlaGluGluAlaArgGlnCysLysGluGlnIleValArgLeuPheAsnAsn 271
 DB 865 GCAACAGTTGGGAAAGCTTCTATGCTTAAGAAAAATTAAGATGTCGCTTTTCATATAT 924
 QY 272 ProArgIleLysIleMetTyrSerAsnAspGluLeuProProGluGln---AspAspThr 290
 DB 925 CAGAGAAATTCAAATCTTTCTCGAACAGTAGCGTGGCCCAATTAATTAATGATATATCA 984
 QY 291 SerPheTyrSerGlyMetLysArgSerArgThrAspMetPheAsnAsnAspProSerCys 310
 DB 985 ACAGCAGTTTCTGATTTCCAAATCA-----GAGATGTATATGATGATGATGATGAT 1038
 QY 311 ValSerSerProHisSerThrGlyIleProGlySerMetArgProLeuArgGlyTyrAsn 330
 DB 1039 GCGGCTTGTGATTA-----TTGATCTCTCGACGGGAAAGATCA 1077
 QY 331 GluArgSerTyrAsnGlyAlaGluTyrAsnAspValValGlyLysGluProAsnTrpArg 350
 DB 1078 AGATAC-----TTTGAGTACAGATGATGTC----- 1101
 QY 351 ArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuPro 370
 DB 1102 -----CCTGTTCTGTGTGATTTCTTCA 1125
 QY 371 SerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTyrGluGlyTyr 390
 DB 1126 TCACCGGAATCGGA-----AATCCTCTCTTAATGAGCATGCTGCTCAGAGTACATTC 1179
 QY 391 AspProAlaGlnLeuValArgGluSerLysArgThrArgArgAspGlySerValAsp--- 409
 DB 1180 GATCA-----AGAGAACCAAGAGTGTGGTGTGATGCTGACGCTGACCT 1227
 QY 410 -----GlyPheThrProMetGlyValAspGluArgSerPhe 421
 DB 1228 TATGATACAGGACGAGTAGAGAGGCTTTATTTCTGTGGTACTCTCAGCGGAGAGT 1287
 QY 422 GlyArgGlySerValAlaAlaArgPro-----IleArgGlyProProAspSer--- 437
 DB 1288 GCACGTTCTGAGAGAGCTCAAGCTCTGATTCGATTCGATTCAGTCAAGTCAAGTCAAGT 1347
 QY 438 -----AspHisIleTyrArgGlyMetIleAlaLysGlyGlyThrProValCysCys 454
 DB 1348 TCATACCTTGAGCACTTGTGGCGTGTGATGCAATTCGCAAGGTGATCTCTCTCTGAT 1407
 QY 455 AlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluValValAsnCys 474
 DB 1408 CTTACACATATTCACATG-----TTTGACAGACCGAGATGTTGTTATTTGT 1452
 QY 475 SerAlaArgThrAspLeuAsnMetLeuAlaLysIleTyrAlaValAlaIleGlyCysGln 494
 DB 1453 TCCCTGAACTGAGACTAGATATGCTGCAAGCATTAAGACGCTTCAGGGTTGAT 1512
 QY 495 IleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArg 514
 DB 1513 ATCGTCTTCTTCTGCAAGATGAGATGATGATGATGATGATGATGATGATGATGATGAT 1572
 QY 515 TyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspAspGlyThrThrLeuPhe 534
 DB 1573 TACTTAGGCTCAAAAAGTCCGGGAGGGGTTGTAAGGTTGATGGGGGATCTTATTTT 1632
 QY 535 LeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeu 554
 DB 1633 TTGCTCCACACTTCCGATTTTTCAGAAATGTTTGAAGTATGATGATGATGATGATGAT 1692
 QY 555 TyrGlyValValLeuLysLeuPro-----ProProAlaValProVal 568
 DB 1693 TAGGTTAGTATTAACATTTCCGCAAAATGTCTGCTGCTCTCTCTCTCTCTCTCTCT 1752
 QY 569 ThrAlaSerTyrArgGlnGlu-----SerGlnSerAsnProLeuHisIleTyrMetAsp 585

Db 1753 CCTGCTGTCAAAAGGCACACTAAGTGCACGAGATTCACAACT-----TACTAGAT 1806
Qy 586 GlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuThrProAlaGluAsnIyr 605
Db 1807 GAA---AGGGAATTCCTTGTGCAGAGAGAGTATGATGATTCAGGCTAGCAACAATTCAC 1863
Qy 606 IleArgGlyAlaProGluHisLeuThrAla----- 615
Db 1864 CATGAGATGCT---GATCACCGGTGATCTTTCAGTGAAGATTCAATTCATCAATTGGGG 1920
Qy 616 -----AlaSerIleProSerValSerGluPro-----LeuArg 626
Db 1921 CAATTCCTTGACAGACTCGGGTGAAGAGCAAGTACTGAGCCAAACCTTGGCCGT 1980
Qy 627 IleProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuAlaThr 646
Db 1981 ATCCCTTCAAAACCCAGAGACTACAGTACAGCCTTCACTCCAGCCTGATATGATAGACACT 2040
Qy 647 LeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHisGlnProMet 666
Db 2041 TTAGCTAAACTTTTGGCAAGTGGGAGTCATCAGCTCTGTGTCATGAGCAGCTTCCCTCG 2100
Qy 667 SerGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeuIyrAsnGlyGlu 686
Db 2101 AGC-----TCAACAGACCAACCTGCATTAACCCAAATGAT-----GATGATCT 2145
Qy 687 AlaProSerGlnAlaThrIlyAsnGlyProGlnThrValHisAspAlaSerAsnGlnSer 706
Db 2146 ACTCTTCAAAAGTATAGAGA-----CTGAAATCAAGCAATGAGCTTCTACTTCATCC 2199
Qy 707 PheGlnGlnIlyrGlyAsnGlnIlyrThrProAlaGlyGlnLeuProProProSerArg 726
Db 2200 TTAGAGCAAGTCCGATAT--TTCCAGCATTCAGACAA-----CAGTTCACCAAA 2247
Qy 727 TyrProProAlaSerAsnAsnProAsnIlyrThrSerGlyMetValHisGlyAsnMetGln 746
Db 2248 CAAGCTGAGACATTCATCTGCCAAACTAT-----GGGAAGCTTGGCAGGTGCACAGAG 2301
Qy 747 TyrGlnSerGln-----SerValAsnMetProGlnLeuSer 758
Db 2302 CACCCCAACCAATCATGATGCTTCAATCTCAATATCTCAATATCTTCAATCTTCAATCT 2361
Qy 759 ProLeuProAsnMetProHisAsnAsnIlyrSerMetIlyrThrGlnIlyrSerSerAsnHis 778
Db 2362 CCACTCTTCAAGCTGTCACCAAGTCTGTATATTCATCGCAAGTTCGACACAGTTCG 2421
Qy 779 ProValSerGlnProMetValGlnGlnIlyrGlnProGluAla---SerMetProAsnGln 797
Db 2422 CCT--ACGCAGATGATGACAGCAAGTATCAACAGAGCATATATATGATGACTCAAGC 2478
Qy 798 AsnIlyrGlyProIleProSerIlyrGlnGlnAlaAsnPheHisGly----- 812
Db 2479 AATTATGATCAATTTGGCAGACAGTATGAGATTCATTCCTTCAAGGCGACCATCAACAAT 2538
Qy 813 ValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAlaAlaMetGlnPro 832
Db 2539 GTTGCACTTCCTGCTCAGCAAGCTCTGTGCTGCGCAATTCCTCTGCTATGACAGCT 2598
Qy 833 ProAlaAspIlyrAlaAsnLeuGluProGlnAsnGlnAlaLeuArgLeuGlnProMetIle 852
Db 2599 CTTGCT-----GCCGCAACAAGCTCTCTGT 2622
Qy 853 SerGlyAspGlyGlnIlyrThrAspGlyGluValAspIlyrAsnGlnIlyrGlnSer 872
Db 2623 GCT-----GCACAAAGCTTCTGTCTGAC--GAGGCAAGAGCAAGAAAGTATCAGGCG 2673
Qy 873 ThrLeuGlnPheAlaAlaAsnLeuLeuGlnIlyrGlnGlnIlyrGlnGlnIlyrGlnSer 892
Db 2674 ACTCTTCAAGTGGCTCAGCGCTGTGGTCACTACCAAGAAAGCTGGAATCAACCT 2733
Qy 893 SerGly-----ThrProAlaGly 898

Db 2734 TAAGGCTATCTTGAAATCAACCTTCAGGC 2763
RESULT 6
US-10-437-963-79215
; Sequence 79215, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 79215
; LENGTH: 3315
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_78947C.1
; US-10-437-963-79215
Alignment Scores:
Pred. No.: 1,9e-62 Length: 3315
Score: 802.00 Matches: 298
Percent Similarity: 37.298 Conservative: 120
Best Local Similarity: 26.584 Mismatches: 325
Query Match: 16,704 Indels: 378
Gaps: 45
US-09-920-705-3 (1-901) x US-10-437-963-79215 (1-3315)
Qy 7 ProPheArgAlaAspAspSerGlyPhe-----GlnSerAsnAsnLeuTrpVal 22
Db 121 CGCGCGGCGGCTCTCTCCGCGTGGGGGCGGACCCGCGCGGACCTTGGGGTG 180
Qy 23 GlySerLeuThrProGluThrThrGluSerAspLeuThrGluLeuPheGlyArgIlyr 42
Db 181 GCGAGCTCTCCCGCGCGCTCCCGCGCGCGGCGGACCTTGGAGCTCTTCCGCGTGGCG 240
Qy 43 AspIleAspArgIleThrValIlyrSerSerArgGlyPheAlaPheIleIlyrIyrArgHis 62
Db 241 GACGTGAGAGGCACTTCCCGGTGACCCCGCGGAGCTTGGCTTGTGACGTTCCGCGCG 300
Qy 63 ValGluGluAlaValAlaAlaIlyrGluAlaLeuGlnIlyrAlaAsnLeuAsnGlySerGln 82
Db 301 GAGGAGGAGCGCGTGGCGCGCGGTGCGGAGCGAGGATTCACCTCGCGCGCGCGCC 360
Qy 83 IleIlyrIleGluIlyrAlaArgProAlaIlyrProCysIlyrSerLeuTrpValGlyIlyle 102
Db 361 ATTAGATGAGATTTTCCAG-----GGGATTA 390
Qy 103 GlyProAsnValSerIlyrAspAspLeuGluGluIlyrPheSerIlyrAspGlyIlyrIleGlu 122
Db 391 GGTTCAGTACGCTCTTATGATGAC-----AGATTCACCAACATGCT 432
Qy 123 AspPheArgPheLeuArgGluArgIlyrThrAlaPheIleAspIlyrIlyrIlyrMetAsp 142
Db 433 GATCAAAAGCGTTTACTGACGAGGA----- 459
Qy 143 AlaLeuGlnAlaIlyrSerMetAsnGlyIlyrProMetGlyIlyrSerPheLeuArgVala 162
Db 459 ----- 459
Qy 163 PheLeuArgSerGlnAlaProIlyrIlyrGluGlnIlyrIlyrIlyrIlyrIlyrIlyrIlyr 182

Db 460 -----AGGATGACGACATCAAGTCTGAAAA-----TCAGCTGAT 495
QY 183 GlysMetAsnHisIysProGlnTyrProHisSerTyrGluAspPheVal 202
Db 496 AAATCCAAAAGAACGCCA-----GCA 519
QY 203 GlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAspGlu 222
Db 520 GAACCTAGTAGAGTATTATGATAGGTTTCTGTTGGTGTGAAGTA-----GATGAG 573
QY 223 GlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyrPro 242
Db 574 GCAACTCTCTGGGAGAGCTTTTCACTTTTGTGAGGTGTCTCAAGATACTACATTCCCA 633
QY 243 SerArgAsnPheAlaLeuValGluPheArgSerAlaGluGlnAlaArgInCysLysGlu 262
Db 634 GGGCGTACTTATGATCTTGTTCAGTACACTACTATTGTCAGCGGCAATGCGGAGGAGAA 693
QY 263 GlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetLysSer 279
Db 694 ACACTCAGGGAATAATTTTTCATAAACCTCCGAGTTAGCATTTGCTTCTCGAGTGCAC 753
QY 280 -----Asn 280
Db 754 AGTGTTCAGCAGAAATTTGGAAAAAGTCTTAGATGCCCAATATCCCCCATTTAAAC 813
QY 281 AspGluLeuProPro-----GluGlnAspAspThrSerPheTyrSerGlyMetLys 297
Db 814 TCTAGTGTTAGACCTTATTTCCAGGAGCAGAAATTTGAAAGATTT----- 858
QY 298 ArgSerArgThrAspMetPheAsnAsnAspPro-----SerCysValSerSerProHis 315
Db 859 ---CCTAGGGCTAGGCTTTGTAGTCTCTCAAGAGATATGATACATCTCCACAT 915
QY 316 SerThrGlyIleProGlySerMetLysProLeuArgGlyThrAsnGluArgSerTyrAsn 335
Db 916 TAT-----GGCCCTAAGAGACTTCTTGA----- 939
QY 336 GlyAlaGluTyrAsnAspValValGlyLysGluProAsnTrpArgArgProSerAlaAsn 355
Db 940 -----GATCATGATATATGTTGGTTTCAGCAGAGATATATATTGCGA----- 981
QY 356 GlyThrGlyIleLeuProSerProThrGlyProGlyIleLeuProSerPro----- 372
Db 982 -----TATGACCTCGAGTAGAGCTGATCTGATCTAT 1017
QY 373 -----AlaGlnGlyThrArgAlaGlyProMetArgSerAsnProAspSerTrp 387
Db 1018 TTGGAACCTTTAGATACAGAGGCTCGGTCCAGAAAGAGATGTCTAGAGACCATAT 1077
QY 388 GluGlyTyr----- 390
Db 1078 GAACAGCATAGGCGTAGCCCTGCTGTGATGACCATGGCACAACATTCCATTCCAGCGA 1137
QY 391 -----AspProAlaGlnLeuValArgGlu----- 398
Db 1138 TCTCAGGAGCCTTACCATTAAGAGATCTCGGTATGCTAGGAGAGATCCATACCATTT 1197
QY 399 SerLysArgThrArgArgAspGlySerValAspGlyPheThrProMetGlyValAspGlu 418
Db 1198 TCAAGAAATTTGAGAGCTGTGTAGAGCATGACTCTGAACTTCCT--GAATACCTCTTC 1254
QY 419 ArgSerPheGlyArgGlySerValAlaAla-----ArgProIleArgGlyPro 434
Db 1255 TCTGATTTGATCGAGGGAAGTGTGCTGTGCTACCCAGAGAGGCCCTTCTATGTGTG 1314
QY 435 ProAspSerAsp----- 438
Db 1315 CCAAGTATGACATACACCCAGAGGCTATCAACTGTCTATGATGATGAATCAT 1374
QY 439 -----HisIle----- 440
Db 1375 GTTGATCTTTTAGAAATCCAACTCACTTGTATGATAGCATATACAGAGGCGATGCACAG 1434

QY 440 ----- 440
Db 1435 GACAGCTTTCTTAGCATGTAGAGTGAAGATCACTCTGTAAATACATGACCCTT 1494
QY 441 -----TrpArgGlyMetIleAlaLysGlyGlyThrProValCysCys 454
Db 1495 CTCAGAGAAATGGAATGGATGGTTCATATAGCAAGGAGGACACCAATTTGCGCA 1554
QY 455 AlaArgCysValProMetGlyLysGlyIleGluThrLysLeuProGluValAlaAsnCys 474
Db 1555 GCGCATGCTCTCCCTGTGGGAAAGGTCTTAACCTCATGCTGCCGGAATTTTGGATTGC 1614
QY 475 SerAlaArgThrAspLeuAsnMetLeuAlaLysHisTyrAlaValAlaIleGlyCysGlu 494
Db 1615 ACTGTAGAGACAGCTGAGATGCTCTTAAGCATTTTTCAAAGCTGCCAGCAGCTGC 1674
QY 495 IleValPhePheValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArg 514
Db 1675 GGGGTGTTTGTTCAGAAATATGCTGTACATGGCAGCTTATATGAATTCATGAT 1734
QY 515 TyrLeuSerSerLysAspArgAlaGlyValAlaLysLeuAspArgLysThrLeuPhe 534
Db 1735 TACCTGGTGTAAAGCAGCGTGCAGAGTGTGTAACTTGAGAAAGGACGAGCTTATTT 1794
QY 535 LeuValProProSerAspPheLeuThrAspValLeuGlnValThrArgGlnGluArgLeu 554
Db 1795 CTGTTCACCTCAGACTTCTTGAAACAGTACTGAGGTTCCAGGTTAAAGTCAGCATA 1854
QY 555 TyrGlyValIleLeuLysLeuProProProAlaValProValThrAlaSerTyrArgGln 574
Db 1855 TCTGAGTCAATTTCTGAATTTGAGCAGATGCAGATCAGAAAGTTTCTGCCCACTGCATA 1914
QY 575 GluSer----- 576
Db 1915 CCAGAAACATTTGAGTCAATTTGAAACCATGATGTTGCTCATGAGATCTAGATGCA 1974
QY 577 -----GlnSerAlaPro-----LeuHisTyrMetAspGlnAlaAsp 589
Db 1975 TTGAGAAATCAACCCACAGATATGAGCCACTCTCCAGGTTCAAGATTATCTCGCG 2034
QY 590 ---SerProAlaAsn-----AlaSerHisSerLeuTyrProProArgGluAsnTyr 605
Db 2035 TTGTGCGCTGAGACTATTAATCCAGCAAGTGCACATTTGTTCCGCTTACAG----- 2088
QY 606 IleArgGlyAlaProGluHisLeuThrAla----- 615
Db 2089 TTGGAATGCTCTTCATATCTGAATCTGAATTAAGTCAATCAAAAGCATCCACTGAC 2148
QY 616 -----AlaSerLysProSerValSerGluProLeuArgIleProAsn--- 629
Db 2149 TCCACAGGAGATAGCATGACACAGCAGCAGCAACCCAGATGTATTTGCCCTCAGAG 2208
QY 630 -----AsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeu 643
Db 2209 TGTGACATACATTTCAATTCACAGTCCAGGTTCTGCA-----AATTG 2253
QY 644 LeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAlaAlaProGluSerHis 663
Db 2254 AATTATTTGGTGAAGTGGATCCCATATACATCACTGATAGCACCAGAGGCATAC 2313
QY 664 Gln-----ProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsnGly 681
Db 2314 TCATTGCTCTCAAGAGATACAAA-----GTAGTCA----- 2349
QY 682 LeuTyrAsnGlyGluAlaProSerGlnAlaTrpLysArgLysProGlnThrValHisAsp 701
Db 2350 -----TCAGGATGATGACCA-----GTTGACATGAG 2376
QY 702 AlaSerAsnGlnSerPhe-----GlnGlnTyrGlyAsnGlnTyrThrProAlaGlyGln 719
Db 2377 GCATCAAAACATGTCTTACCTCCATGCACACTGCATACACAGAGTATGATGACCTGCA 2436

```

Qy 720 LeuProProPro---ProSerArgTyProPro----- 729
    |||||
Db 2437 CAACCTTCACTCTCCCATTTATGCTTCCACACAGACAATTGCACATTGGCCACTCTT 2496
Qy 730 -----AlaSerAsnPro-----AsnTyThrSerGly 739
    |||||
Db 2497 CTTCGACAAACAAACCAACGAGAAAGAGCTGTGACAGCTTAAACAAAGATCAACGA 2556
Qy 740 MetVal-----HisGlyAsn-----MetGlnTyrglnSerGlnSerValAsn 753
    |||||
Db 2557 TTCAATGGAATCAATCCAGATTCCTTCAATGATGACACAGCTAGGTTCTATC--- 2613
Qy 754 MetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyrsMetTyThrGln 773
    |||||
Db 2614 ---CCTTCCAAACCTATTGCACTGCTCTCG-----CCA 2646
Qy 774 GlySerSerAsnHisProValSerGlnProMetValGln----- 786
    |||||
Db 2647 TCTGCATCAGAGTTACAGGTTTCAAGGCCACACAGTTCAAGTTCAAGTCCAGCCAAATCA 2706
Qy 787 -----GlnTyrglnProGlnAlaSerMetProAsnGlnAsnTyrglyProIleProSer 804
    |||||
Db 2707 TCCATTATGCAATACCGAATGCTCTATGCTTCTCAACAACCTTGCCTTTACCCCT 2766
Qy 805 TyrGlnGlnAlaAsnPhenHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSer 824
    |||||
Db 2767 ATGCATCTCTGGGA-----AATCAGCTCATTTTCCATGCTTTG 2808
Qy 825 GlnPheGlnAlaAlaMetGlnProProAlaAspIysAlaAsnLeuGlnProGlnAsnGln 844
    |||||
Db 2809 AGATCGTTCGCTCCCTCTCTTCCGAGAGCCCTCCACCCCTTAGGCAGCACATCAACT 2868
Qy 845 AlaLeuArgLeuGlnProMetIleSerGlyAspGlyGln-----Gly 858
    |||||
Db 2869 GCTCTACAGACGCACTGCACTTCCCTGACACCAAGACTAGCCAGCAACGCTCTGCT 2928
Qy 859 ThrThrAspGlyGlnValAlaAspIysAsnGlnArgTyrglnSerThrLeuGlnPheAlaAla 878
    |||||
Db 2929 CAGGAACACATCAGCAGAGACCTCAAAAGGCTTCAAGCAACATTGCAATTGGCAGCA 2988
Qy 879 AsnLeuLeuGlnGlnIleGlnGlnIleGlnGlnIleGlnGlnIleGlnGlnIleGlnGln 898
    |||||
Db 2989 ACCTTACTTACGACGATACGACATCACTTAA-----CCTGGTGGC 3030
Qy 899 Gln 899
    |||
Db 3031 CAG 3033

RESULT 7
US-10-424-599-59704
; Sequence 59704, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5323)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 59704
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_24924C.1
US-10-424-599-59704
Alignment Scores:
Pred. No.: 5.64e-51 Length: 1101

```

```

Score: 669.00 Matches: 169
Percent Similarity: 54.57% Conservative: 40
Best Local Similarity: 44.13% Mismatches: 77
Query Match: 13.93% Indels: 97
DB: 17 Gaps: 15

US-09-920-705-3 (1-901) x US-10-424-599-59704 (1-1101)
Qy 282 GlnLeuProProGlnGlnAspThrSerPheTySerGlyMetIysArgSerArgThr 301
    |||||
Db 54 CAGTTTACCACT-----AGTGGTATTT----- 74
Qy 302 AspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIleProGly 321
    |||||
Db 75 ---ATGGAACCAACATACCAATGCACTTTGGTCTCAGCT---GGTGT----- 122
Qy 322 SerMetArgProLeuArgIleThrAsnGlnArgSerTyAsnGlyAlaGlyTyraAsnAsp 341
    |||||
Db 123 -----GAACTGTTATTTCTGGCCCTGAGTTTAATGAG 155
Qy 342 Val-----ValGlyIysGluProAsnTrp 349
    |||||
Db 156 ATTAATGCGCTCCACAAATTTACAGATGAGCTCCAGAGTAGCATGGGTCCAAACTGG 215
Qy 350 ArgArgProSerAlaAsnGlyThrGlyIleLeuProSerProThrGlyProGlyIleLeu 369
    |||||
Db 216 AAAAG-----CAATCTCTCTGCACACGAGGATGCTT 248
Qy 370 ProSerProAlaGlnGlyThrArgArgProMetArgSerAsnProAspSerTrpGlyGly 389
    |||||
Db 249 TCTTCTCTGACACCGGCTGAGCTTCTTCAAGATCAACCTCGGGTGCATGGATGCTG 308
Qy 390 TyrAsp-----ProAlaGlnLeuVal---ArgGlnSerIysArgThr 402
    |||||
Db 309 CTTCGACATAAACCCTATCATTCAGAGAGACACCTTGCTCCAAAGATTTCAAGCTTC 368
Qy 402 LArgArgAspGlySerVal-----AspGlyPheThrProMet----- 414
    |||||
Db 369 AAGGATGATGAGACCTTGGCTGTTGATGAAAGCCCAATTTCTTTAGAGAAATATGATGA 428
Qy 415 -----GlyValAspGlnArgSerPheGlyArgGlySer 425
    |||||
Db 429 TCGGGGGTTAGCTCTAGACAGACATATGGAGATTAACCAAGCATTTATGATGAGAGTGCTT 468
Qy 425 I-----ValAlaAlaArgProIle 431
    |||||
Db 489 TGGTCATATGTAACATTCAGAGAAAGATCACTTGGTCCAGTTAGCTACAGAGATTAC 548
Qy 431 eaArgGly-----ProProAspSerAspHisIleTrpArgGlyMetIleAl 446
    |||||
Db 549 AGCTGAGATGATGATGTCGCCCAACCTGATATGATCAATATTTGCGTGGAGATTATGTC 608
Qy 446 alysglyGlyThrProValCysCysAlaArgCysValProMetGlyIysGlyIleGluThr 466
    |||||
Db 609 AAAAGGGGAACCTCTGTTTGCCTGCTAGATGTGACCAATTTGGGAAAGGATTTGGGAC 668
Qy 466 rlyLeuProGlnValValAsnCysSerAlaArgThrAspLeuAsnMetLeuAlaIysHis 486
    |||||
Db 669 TGAGCTTCCGATGATGTTGATGCTCAGCTAGAGCGGATGATGATATCTCAAAACAA 728
Qy 486 sTyraAlaValAlaIleGlyCysGluIleValIlePhePheValProAspArgGlnGluAsp 506
    |||||
Db 729 TTAATGCTGATGCAATTTGTTGATATGTTTCTTTCGCTGAGATGATGACAAACATTT 788
Qy 506 eaIAserTyThrGlnPheLeuArgTyrlleuSerSerIysAspArgAlaGlyValAlaIys 526
    |||||
Db 789 TGCTTCAATACCGAATTTCTTCTGCTATCTTAGTGGAAAAATCGTGGTGTTCACAA 848
Qy 526 sleuAspAspGlyThrThrleuPheLeuValProProSerAspPheLeuThrAspValle 546
    |||||
Db 849 ATTTGTTTAACACACCACTTATTTCTTGGTGCCTTCTGATTTTCTCACAAAGAGTTT 908
Qy 546 uGlnValThrArgGlnGlnIleArgLeuTyrglyValValIleuLysLeuProProAlaVal 566

```

```
Db 909 GAAAGTACTGGAGCTGAGCGCTGATGAGTGCTTTAAGTTCCACCA-----GT 962
Qy 566 lProvalThrxlaserTyhrargingluserGlnSerAsn-----Proleuhistryme 584
Db 963 GCCAAGTAGTGCACCTTATGCAACACCATCATTTCCGTGACCAACTACTAGATATAT 1022
Qy 584 lAspGlnAlaArgAspSerProAlaAsnAlaSerHisSerLeuTyProProArgGluAs 604
Db 1023 GCAGCAG-----AATCTCTCTTCGACACTGAATATGCTTGTATCTCTGTAAGAGGA 1076
Qy 604 nTyrlle 606
Db 1077 ACATATT 1083

RESULT 8
US-09-920-705-6
; Sequence 6, Application US/09920705
; Publication No. US2003079252A1
; GENERAL INFORMATION:
; APPLICANT: Amasino, Richard M.
; APPLICANT: Schomburg, Fritz M.
; APPLICANT: Michaels, Scott D.
; APPLICANT: Patton, David
; TITLE OF INVENTION: Floral Induction Gene
; FILE REFERENCE: 960296.97214
; CURRENT APPLICATION NUMBER: US/09/920,705
; CURRENT FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 3715
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: FPA promoter plus intron
US-09-920-705-6

Alignment Scores:
Pred. No.: 1,09e-48 length: 3715
Score: 651.50 Matches: 171
Percent Similarity: 27.23% Conservative: 0
Best Local Similarity: 27.23% Mismatches: 0
Query Match: 13.57% Indels: 457
DB: 10 Gaps: 1

US-09-920-705-3 (1-901) x US-09-920-705-6 (1-3715)
Qy 1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
Db 1832 ATGGCGTTATCTATGAAGCATTCAGAGCCGATGATTCGCGTTTCCAGTCAAAACATCTT 1891
Qy 21 TrpValGlySerLeuThrProGluThrThrGlnSerAspLeuThrGluLeuPheGlyArg 40
Db 1892 TGGGTGCTAGCTTAACGCGGAGACGACGAGCTGATCTGACCCAGTGTGTTGAGACA 1951
Qy 41 TyGlyAspIleAspArgIleThrValTySerSerArgGlyPheAlaPheIleTyTyP 60
Db 1952 TAGGGGATATTGATAGATTCACGCTGATCTTCCAGAGCTTTGCGTTTATATATC 2011
Qy 61 ArgHisValGlnGluAlaValAlaAlaValGlnAlaLeuGlnGlyValaAsnLeuGngly 80
Db 2012 AGACATGTGGAGGAGACAGTCGCGACGCAAGAGCTCTTCAAGAGCAAAATTTGAATGA 2071
Qy 81 SerGlnIleLysIleGlnTyPAlaArgPro----- 90
Db 2072 AGTCAAAATTAAGTCAAAATTCGACGCGGTTGTCTTATCTATATCTTGCTTGTTC 2131
Qy 90 ----- 90
Db 2132 TCTAATTGATGTCTTTTGTCAAGATTATATCTTTTGGGAATTATAGTCCAGGT 2191
Qy 90 ----- 90
```

```
Db 2192 TCACAAACTTTGATGATGCTGTTTATGTCACAAAATTTCTTGATCTGTTTTTTT 2251
Qy 90 ----- 90
Db 2252 TTCCTATCGTACAAATCAAGTCGAAACCTAGTTTTTTTCTATATATAGTGTTAGCTT 2311
Qy 90 ----- 90
Db 2312 AAGGCGAAACCTGATCCGATCGAAACGTCCTTCTCAAAATTACTTTGGTTATATGAACT 2371
Qy 90 ----- 90
Db 2372 CCGGCAAGCCAAACCAAGAGAGCTTCGAAAAATTTGATGTAAAGCATATATATCTC 2431
Qy 90 ----- 90
Db 2432 TTAGCAATGAGCTCTGCAAAAGAAATACATCAACACATGTTTACTCTGTTTATGCGAA 2491
Qy 90 ----- 90
Db 2492 GAAGTTTAATCCGATGTCGTTTATCTGCAAACTGTATCCGCGTTAGACGTCGATT 2551
Qy 90 ----- 90
Db 2552 TCAGTTGCGCTTCAGATGTTAAATCTCACAGCTTGAGATGATATGTTGCGTACTCA 2611
Qy 90 ----- 90
Db 2612 CTCGAAATTCGCAATGTGTGAAATTTGGAAGGAGCAAACTCTATCATCGGCCAAACAGA 2671
Qy 90 ----- 90
Db 2672 TAAGAAATTTGGAGTTTAAAGTTCAAGTTCTCGCAAAATCAAAACCGGTGAGAAAT 2731
Qy 90 ----- 90
Db 2732 TTGTCTATGGCGGTACGTAGATATCATATACATCTGCGAAACGAAATTTCTGGCAA 2791
Qy 90 ----- 90
Db 2792 CCGATTCGTCTTACTTACTAGTATGCTGACATATCATATCATATGATGAGAGATT 2851
Qy 90 ----- 90
Db 2852 GGGCTTGTGTTGGCTCGATGCTTCAAGAAACAAATTAAGTGTGTTAAGGCGGCTAAC 2911
Qy 90 ----- 90
Db 2912 TCTACCAATCAGAAACGCTTATTCGAGAAACCATGTTTGTTCCTCAATTCATCCCTA 2971
Qy 90 ----- 90
Db 2972 CGTACATCTGGGCTTTCCTCATTTGATGCTTCTAGAAAGTTGACTTCTTCAACATTC 3031
Qy 90 ----- 90
Db 3032 TGGGATTGTACAGTTGACGTGACAGACAAATATTCATTGTCAGGACGACCTAAATTT 3091
Qy 90 ----- 90
Db 3092 CAATGCTTGGCGGCTTGAATACTTAATGTGAGATGTTCAAGGATGGTTTATTTGGGAC 3151
Qy 90 ----- 90
Db 3152 TCAGGCATGATTAAGACAGAAATGTTTGCAGAAAGTATTAATGCTTCCGCGCTTGG 3211
Qy 90 ----- 90
Db 3212 GGGACTATATCTTCTCTACAGACAAATCTGATGTTTAGGCAAGAGAAACTATTAACA 3271
Qy 90 ----- 90
```

Db 3272 CTGATTTAATAGAGAAAGAGATGTTCTCTGGCGCAATTTATTTGTTAGAA 3331
Qy 90 ----- 90
Db 3332 AGCAATTGATATGATGGTGTCTAGTAGTTGAAATTAATTAAGTAGTTGTGTGTT 3391
Qy 90 ----- 90
Db 3392 AGTTTCTTGATGTTTGTATGCTTATTTCTTGCAACCTATCTGGGTAGTATGCCCTTCT 3451
Qy 91 ----- 3451
Db 3452 ATGCACCTTTCTTGTTCAGAGCAAACTTGTAAAGTCTATGGTGGTGGATCGGC 3511
Qy 104 ProbenValSerlyAspAspLeuGluGluPheSerlyPheGlyLysIleGluAsp 123
Db 3512 CCTATGCTCTCCAAAGATGACCTGAGAGAAAGTTTCAGCAAGTTGGGAAATGAGAGAT 3571
Qy 124 PheArgPheLeuArgGlyArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAla 143
Db 3572 TTTAGGTTTCTCAAGAAAGCAAGACGCTTTCAITGTTATTAAGATGATGATGCT 3631
Qy 144 LeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPhe 163
Db 3632 TTACAGGCTAAGACATGAATGAAAGCCTATGGGTGTAGCTTTTGGCGTTGATTTT 3691
Qy 164 LeuArgSerGlnAlaProLysLys 171
Db 3692 CTCGGTCAACAGCGCCAAAAA 3715

RESULT 9
US-10-424-599-65572
; Sequence 65572, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 65572
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(801)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_30226C.1
US-10-424-599-65572

Alignment Scores:
Pred. No.: 1,09e-35 Length: 801
Score: 499.00 Matches: 105
Percent Similarity: 65.55% Conservative: 32
Best Local Similarity: 50.24% Mismatches: 63
Query Match: 10.39% Indels: 9
DB: 17 Gaps: 3

US-09-920-705-3 (1-901) x US-10-424-599-65572 (1-801)
Qy 6 LysProPheArgAlaAspAspSerGlyPheGlnSerAsnMetLeuTyrValGlySerLeu 25
Db 165 AAATCCGTGAAAGCGAGAGTGGGAAACCCCAACAACCTCTGGGTGGGAAATCTG 224
Qy 26 ThrProGluThrThiGluSerAspLeuThrGluLeuPheGlyArgTyrGlyAspIleAsp 45
|||||

Db 225 CCAGCGAGGCTCAACAGATTCAGATTCAGAACTATGCTCCGTAAGGTTCTCTGAC 284
Qy 46 ArgIleThrValLysSerSerArgGlyPheAlaPheIleTyrTyrArgHisValGluGlu 65
Db 285 TCACCTCATCTCTCACTCTCTCCGACCTTCGCTTTCGTTGTTCAAGGCCATCGAAGAC 344
Qy 66 AlaValAlaIleValGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGlnIleLysIle 85
Db 345 GCTAAGGACCGAAATCAACCTTGCAAGCGCGCTGTGCGCGGTTTCAATCAAGATC 404
Qy 86 GluTyrAlaArgProAlaLysProCysLysSerLeuTyrValGlyLysIleGlyProAsn 105
Db 405 GAGTTTCCAGACCGGAAAGCCGTGCAACATTTGGTGGTGGTGGTTTATGTCCTACT 464
Qy 106 ValSerLysAspAspLeuGluGluGluPheSerLysPheGlyLysIleGluAspPheArg 125
Db 465 GTTGCAAGGAGAAATTTGGAAGCGGAATTCGCAAAATTCGTTAAGATTTGAAG 524
Qy 126 PheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMetAspAspAlaLeuGln 145
Db 525 TTTCTTATAGACCGCGGACCTGCTGTGTGAATTTCTCAATCTGAGATGCTGCTCGG 584
Qy 146 Ala---LysSerMetAsnGlyLysProMetGlyGlySerPheLeuArgValAspPheLeu 164
Db 585 GCGATGAAGTCAATGAACGGGAAGCGGNTAGTGAAGCCAGATTTGTAGATTTCTT 644
Qy 165 ArgSerGlnAlaProLysLysGluGlnTyrPheGlySerTyrAspAsnArgAsnGlyAsn 184
Db 645 CGATCAACATCTACGAGAAAGAT-----TTTGTGTTGATCATGTGGCGAG 689
Qy 185 MetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysGlyAspValGlnPro 204
Db 690 TTTCAAGGCTAGACCAAG-----CATTAACAACCTTCAATTGGAGGAATTAATCAACC 743
Qy 205 SerLysValLeuTyrPheIleGlyPhe 212
Db 744 AGTAAATAATCTGTGGATGAGGTTT 768

RESULT 10
US-09-920-705-5
; Sequence 5, Application US/09920705
; Publication No. US20030079252A1
; GENERAL INFORMATION:
; APPLICANT: Amadio, Richard M.
; APPLICANT: Schomburg, Fritz M.
; APPLICANT: Michaelis, Scott D.
; APPLICANT: Patton, David
; TITLE OF INVENTION: Floral Induction Gene
; FILE REFERENCE: 960296.97214
; CURRENT APPLICATION NUMBER: US/09/920,705
; CURRENT FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 349
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Portion of FPA coding region
US-09-920-705-5

Alignment Scores:
Pred. No.: 3.68e-33 Length: 349
Score: 466.00 Matches: 92
Percent Similarity: 94.95% Conservative: 2
Best Local Similarity: 92.93% Mismatches: 3
Query Match: 9.71% Indels: 2
DB: 10 Gaps: 1

US-09-920-705-3 (1-901) x US-09-920-705-5 (1-349)
Qy 1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnMetLeu 20
|||||

Db 1306 GGAACACAGTTG-----GATAGCAGCTTGACAGAAAGTCTGATGTGATCCCA 1353
Qy 867 AsnGlnAArgTyGInSerThrLeuGlnPheAla1AsnLeuLeuGlnIleGln 885
Db 1354 CAGAGAGCTGTACACAGAACACACTGACGTGGCGCTGTTCTTCCACCAATCCAG 1410

RESULT 12

US-10-424-599-40095
; Sequence 40095, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 40095
; LENGTH: 1631
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: clone ID: PAT_MRT3847_136205C.1
US-10-424-599-40095

Alignment Scores:

Pred. No.: 1.5e-30 Length: 1631
Score: 447.00 Matches: 155
Percent Similarity: 42.80% Conservative: 59
Best Local Similarity: 31.00% Mismatches: 195
Query Match: 9.31% Indels: 91
DB: 17 Gaps: 17

US-09-920-705-3 (1-901) x US-10-424-599-40095 (1-1631)

Qy 432 ArgGlyProProAspSerAspHisIle-----TrpArgGlyMetIle 445
Db 49 AGATTACACGCTGATCCGATTCGATTCATCTTAACTGAGTGAATGGAAAGAACATT 108
Qy 446 AlAlaGlyGlyThrProValCysCysAlaArgCysValProMetGlyGlyIleGlu 465
Db 109 GCTAAAGCGGAAACCCCTGTTGTTGTCAGCGCTGCTCCCTGGGAAAGTCTAGAT 168
Qy 466 ThrIleuPheProGluValAlaAsnCysSerIleArgThrAspLeuAsnMetLeuAlaLys 485
Db 169 ATGATGTTACTGAGTCTTGATTTGACCTGCAAAAACCTGCTAAGATGCTTTTCAAG 228
Qy 486 HisTyraIaValAlaIleGlyCysGluIleValPhePheValProAspArgGluAsp 505
Db 229 CATTAACACCAAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 288
Qy 506 PheAlaSerTyThrGluPheLeuArgTyIleuSerSerIysAspArgAlaGlyValAla 525
Db 289 ATGGAAGTCATCAATGATTTATCTGATATCTGAGAGAAAGAGCCGCTGACGTTGCC 348
Qy 526 LysIleuPheAspArgTyThrIleuPheLeuValProProSerAspPheLeuThrAspVal 545
Db 349 AAGTTGATGACAAACACCACTTATTTCTGTTCTTCATCAGAGTTCTCAGAGAAAGTA 408
Qy 546 LeuGlnValThrArgGlnGluArgLeuTyGlyValValLeuLysLeuProProProAla 565
Db 409 CTGAAGGTAAGTCTGATAGTATCTGATATCTGATATCTGATATCTGATATCTGAT 466
Qy 566 ValProValThrAlaSerTyf-----ArgGlnIleuSerGlnSerAsnProLeuHisTyf 583
Db 469 TTAATATCATGCTCTGAACACATTCATAAAGAAATGACTAAACAAATTCGCTCTAT 528
Qy 584 MetAspGlnAlaArgAspSerProAlaAsnAlaSerHis-----SerLeuTyfProPro 601

Db 529 AATGAG-----AATATACCTGATCTGAAGCTATCATTTCTCTCA 567
Qy 602 ---ArgGluAsnTyrlleArgGlyAlaProGluHisLeuThrAla----- 615
Db 568 GTACGTGATACCTATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 627
Qy 616 -----AlaSerLysProSerValSerGluProLeuAlaGliePro 628
Db 628 TCATTTCTGGAAATAATTTGAGACAGCTCCATCAGTTCTGATTCATTCAGCTCGT----- 661
Qy 629 AsnAsnAlaIleProGlnAlaGlyValSerLeuThr----- 640
Db 682 -----GCTGCGCTAGCATTCCTGAGTTTCATGATGAAGAAGTCAC 723
Qy 641 -----ProGluLeuAlaThrLeuAlaSerIleLeuProAlaThrSerGlnProAla 658
Db 724 AACTATCTTACCCAGACGCGTACTTCTTCAAAATCTCCAGAACTTCTTAATAAGCT 783
Qy 659 AlaProGluSerHisGlnProMetSerGlyProSerThrValValSerThrAlaHisGln 678
Db 784 CTACCA-----TTGAGCGCATCAAGTGTGCTGAGAACCTTATTCCTGATGAACACCA 837
Qy 679 SerAsnGlyLeuTyraAsnGlyAlaIleProSerGlnAlaIleTrpLysArgGlyProGlnThr 698
Db 838 -----CCTATCATTCCAAGGCGCAGCGAGAT 864
Qy 699 Val-----HisAspAlaSerAsnGlnSerPheGlnIleTyGlyAsnGlnTyf 714
Db 865 GTAAAGCGCATTCACATCCCATCGGAAATTTCTGATTCATTTATGATGAT----- 918
Qy 715 ThrProAlaGlyGlnLeuProProProProSerArgTyf-----Pro 728
Db 919 -----AGCAAGTGTCTTATCCAGATATGAAGATTAGATCCTTATCAGTGCCT 969
Qy 729 ProAlaSerAsnAsnProAsnTyfThrSerGlyMetValHisGlyAsnMetGlnTyfGln 748
Db 970 GTGGAGGCTCTCTCCAGAACAGCTTGACCAATCAGCAACGCTTCTTGACGACGAG 1029
Qy 749 SerGlnSerValAlaAsnMetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyf 768
Db 1030 AGGCATTCGGAGAGGCTTCGATACATCTGCGTGGTGGAT-----CCCAGCAGATTAA 1086
Qy 769 SerMetTyfThrGlnGlySerSerAsnHisProValSerGlnProMetValGlnGlnTyf 788
Db 1087 AGCTTTGGCAGATCCGACACTCATCATCAGATTATCTCA----- 1125
Qy 789 GlnProGluAlaSerMetProAsnGlnAsnTyfGlyProIleProSerTyfGlnGlnAla 808
Db 1126 TCCGACAACTCATTCAGACACCCCGAAATATGCTACAGAAACATGTGGTAATTC 1185
Qy 809 AsnPheHisGlyValThrThrAsnGlnAlaGlnAsnLeuAsnProSerGlnPheGlnAla 828
Db 1186 GATCTCTCAGCATCTGATGGGTCAAAAGTCAAAATGCTACAGTGAACAAACAAACAG 1245
Qy 829 AlaMetGlnProPro-----AlaAspLysAlaAsnLeuGlnProGlnAsnGlnAlaLeu 846
Db 1246 ATGTAAATGTGCTCAGTGTGACAGATAGTTCAAAAGGAAACCAAAAGAGAGGCAAT 1305
Qy 847 ArgLeuGlnProMetIleSerGlyAspGlyGlnGlyThrThrAspGlyGluValAspLys 866
Db 1306 GGAACACAGTTG-----GATACGAGCTTCGAGAGAAATGATCATGATATGATCC 1353
Qy 866 AsnGlnAArgTyGInSerThrLeuGlnPheAla1AsnLeuLeuGlnIleGln 885
Db 1354 ACAGAGAGCTGTACAGCAACACTGACGTGGCGCTGTTCTTCCACCAATCCAG 1411

RESULT 13

US-10-425-114-31541
; Sequence 31541, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 8, 2005, 08:31:28 ; Search time 6977 Seconds

(without alignments)
4915.566 Million cell updates/sec

Title: US-09-920-705-3

Perfect score: 4801
Sequence: 1 MALSMKPRFRDSDGFSQNNL.....LQIQKQKQSSGSGTPAGQGP 901

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool.h/US09920705/runat_07042005_101813_20668/app_query.fasta_1.1095
-DB=EST -OPMT=fastrap -SUFFIX=rcst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09920705_@CGN_1_1_4087_@runat_07042005_101813_20668 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4263.5	88.8	3082	3	CNS0A83E
2	928.5	19.3	767	8	BH552322
3	808.5	16.8	866	9	CL511524
4	751	15.6	724	8	BH552335
5	747	15.6	812	5	BX840244
6	716.5	14.9	2706	3	CNS0A2CE
7	695.5	14.5	602	4	BG543956
8	693	14.4	701	2	BE037904
9	607	12.6	635	1	AU237346

10	561	11.7	416	8	BH851246	BH851246	SAUK_0727
11	537	11.2	582	1	AV829561	AV829561	
12	535	11.1	817	7	CO095058	CO095058	GR_BA171
13	532.5	11.1	800	6	CA766318	CA766318	AF53-RPc
14	532.5	11.1	1105	7	CK208950	CK208950	FGAS02067
15	511.5	10.7	582	7	CF322873	CF322873	HDN_-02-E
16	510	10.6	685	5	BH831648	BH831648	T024803_P
17	507	10.6	412	8	BZ763896	BZ763896	SAUK_1226
18	500	10.4	281	6	CB260919	CB260919	52-E5517-
19	493	10.3	655	2	AM696401	AM696401	NFL03E07S
20	486	10.1	630	4	BG447636	BG447636	NFL18B10S
21	470	9.8	295	9	AL938124	AL938124	Arabidops
22	461	9.6	362	8	BZ288571	BZ288571	SAUK_0219
23	454.5	9.5	624	5	CA032838	CA032838	HX14F14R
24	452	9.4	561	2	BF266387	BF266387	HY_C8A001
25	451	9.4	617	4	BG649121	BG649121	EST510740
26	450.5	9.4	683	3	CK851161	CK851161	11245_Stc
27	450.5	9.4	855	6	CB656009	CB656009	OSJNEC090
28	448	9.3	662	2	BF641487	BF641487	NF064H031
29	441	9.2	421	5	BP647781	BP647781	BP647781
30	430.5	9.0	441	6	CD481691	CD481691	eca01_17m
31	423	8.8	647	5	BQ696544	BQ696544	NXPV_042_
32	418	8.7	542	7	CO910873	CO910873	B03012A1
33	409	8.5	546	7	CN893760	CN893760	010611AAY
34	406.5	8.5	754	6	CA205899	CA205899	SCVPFL107
35	406.5	8.5	793	3	CC634984	CC634984	OCUG295TV
36	403	8.4	493	2	BF639287	BF639287	NF030C091
37	400.5	8.3	814	6	CB633235	CB633235	OSTEB11P
38	399	8.3	278	9	AL938125	AL938125	Arabidops
39	398	8.3	685	4	B1954416	B1954416	HVSEMEM01
40	396	8.2	950	9	CL241131	CL241131	ZMMWB058
41	385	8.0	831	7	CNE21077	CNE21077	GO0105_B3
42	384	8.0	477	7	CF759151	CF759151	DSAF1_41
43	379	7.9	854	9	CG960265	CG960265	MBEGS83TR
44	377	7.9	381	5	BQ697581	BQ697581	NXPV_058
45	376	7.8	666	7	CV092674	CV092674	FAMU_USDA

ALIGNMENTS

RESULT 1
CNS0A83E
LOCUS
DEFINITION
CNS0A83E
Arabidopsis thaliana Full-length cDNA Complete sequence from clone GSU7PH602D02 of Hormone Treated Callus of strain col-0 of Arabidopsis thaliana (thale cress).

ACCESSION
BX820695.1
VERSION
BX820695
KEYWORDS
HTC; GSUT CDNA.

SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE

1 (bases 1 to 3082)
Castelli,V., Aury,J.M., Jallion,O., Wincker,P., Clapet,C., Menard,M., Craud,C., Quetier,F., Scarpelli,C., Schachter,V., Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences: A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
2 (bases 1 to 3082)
Genoscope.
Direct Submission
Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli V., Aury J.M., Jallion O., Wincker P., Menard M., Craud C.,

Schachter V., Weissenbach J., Salanoubat M.

URCV INRA : Clépet C., Caboche M.

Annotation is based on the June 2003 version of the Arabidopsis genome released by MIPS (Munich Information center for Protein Sequences). 5 prime and 3 prime are assembled with Phrap.

http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full

length

<http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis>.

FEATURES

SOURCE

location/Qualifiers

1..3082
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Col-0"
 /db_xref="taxon:3702"
 /clone="GSLTPGH60ZD02"
 /issue_type="Hormone Treated Callus"
 /plasmid="pCMVSPORT_6"
 /complement(1..3082)
 /gene="At2g43410"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 3082
 Score: 4263.50 Matches: 840
 Percent Similarity: 92.74% Conservative: 3
 Best Local Similarity: 92.41% Mismatches: 15
 Query Match: 88.80% Indels: 54
 DB: 3 Gaps: 1

US-09-920-705-3 (1-901) x CMS0A83E (1-3082)

```

QY 1 MetAlaLeuSerMetClyPro--PheArgAlaAspSerGlyPheGlnSerAsnAsnL 20
Db 234 ATGGCGTTATCTATGAGCATGTCAGAGCCGATGATTCGGTTCCATCTCAACATC 233
QY 20 eutrpValGlySerLeuThrProGluThrThrGlySerAspLeuThrGlnLeuPheGlyA 40
Db 294 TTGGGGTCGGTAGCTCAACCGCGAGACGACAGATCAGATCTGACCGAGTTTGGAA 353
QY 40 rGTYrGlyAspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrT 60
Db 354 GATACGCCGATATGATGATGATCAAGATCGATGATCTTCCAGAGCTTTGCTTATATCT 413
QY 60 YrArgHisValGluGluAlaValAlaAlaLysGluAlaLeuGlnGlyAlaAsnLeuAsnG 80
Db 414 ACAAGACATGTGAGAGAGCAGTCGACGCAAGAGGCTTTCAAGAGCAAAATTTGAAG 473
QY 80 lYserGlnIleLysIleGluTyrAlaArgProAlaLysProCysLysSerLeuTTPValG 100
Db 474 GAAGTCAAATTAAGATCGAATACGACGACGGCAAAACCTTGTAAGATCTATGGGTGG 533
QY 100 lYgIlyIleGlyProAsnValSerLysAspAspLeuGluGluGluPheSerLysPheGlyL 120
Db 534 GTGAATCGGCCCTTAATGCTCCAAAGATGACCTGAGAGAGAGGTTCAAGCAAGTTGGGA 593
QY 120 YrIleGluAspPheAspPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluM 140
Db 594 AAATTCAGAGATTTTAGGTTCTCAGAGAACGCAAGACAGCTTCATTGATTATGAGA 653
QY 140 eLAspAspAlaLeuGlnAlaLysSerMetAsnGlyLysProMetGlyGlySerPheLeuA 160
Db 654 TGGATGATGCTTACAGGCTACAGACATGATGAAAGGCTATGGGTGGTGGCTTTTTC 713
QY 160 rGlyAlaAspPheLeuArgSerGlnAlaProLysLysGluGlnTTPAlaGlySerTyrAspA 180
Db 714 GTGTTATTTTCCCGGTCAAGCGCCCAAAAAGAACATGGCTGGCTCTTACAGAA 773
QY 180 snAspArgnGlyAsnMetAsnHisLysProGlnTyrProHisSerTyrGluAspPheLysG 200
Db 774 ACAGAAATGCAATTAAGATCATTAACCCCAAGATTCCTACATATGAAAGACTTTAAAG 833
QY 200 lYAspValGlnProSerLysValLeuTTPilGlyPheProProThrAlaThrGlnCysA 220

```

```

Db 834 GAAATGTCACGCCAAGTAAGTTCTGTGATTTGGTTCCCTCTACTGCTACCAATGCA 893
QY 220 snAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLys 240
Db 894 ATGATGAGCAAAATTCGACAAATGCAATGATATCTCTTTGTGATGATCGAAGGCTAAAA 953
QY 240 eTyrTrpSerArgAsnPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnC 260
Db 954 GTTACCCATCAAGAAATTTTGACATTTGTGAGATTTAGAGCCGCGAGAACTCCGCAT 1013
QY 260 YrLysGluGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerA 280
Db 1014 GCAGAGAAAGCCTTACAGGAGAGGATTTTCAATTAATCTTGAATCAAAATTAATGACTCA 1073
QY 280 snAspGluLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerA 300
Db 1074 ACAGATGATGGCTCCTCGTAGAAGACGATACATGATTTTACTCT----- 1117
QY 300 rGThrAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIleP 320
Db 1117 ----- 1117
QY 320 roGlySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyAlaGlyTyrA 340
Db 1118 -----GAATACA 1124
QY 340 snAspValAlaGlyLysGluProAsnTTPArgArgProSerAlaAsnGlyThrGlyIleL 360
Db 1125 ATGACGTTGTTGTGAAGAGCCAACTGAGAGAGCCATCTGCAATAGAACTGAATAC 1184
QY 360 euProSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgProM 380
Db 1185 TCCCATCTCCACAGACAGCTGGAATCTCCATCTCTCCACAGAGTACAGAGGCCCTTA 1244
QY 380 eLArgSerAsnProAspSerTTPGluGlyTyrAspProAlaGlnLeuValArgGluSerL 400
Db 1245 TGAGGTCAAAACCCGATCTTGGAGAGATGATGATCTGCTCAAGTTGTCAGAGAACTA 1304
QY 400 YrArgThrArgArgAspGlySerValaAspGlyPheThrProMetGlyValaAspGluAs 420
Db 1305 NACGAACCAAGAAATGATGATCAAGTGAAGGTTTATCCNAATGGGTCTCATATGAGAGT 1364
QY 420 ePheGlyArgGlySerValAlaAlaArgProIleArgGlyProProAspSerAspHis 440
Db 1365 CATTTGGTCAGAGTTCAATGCTGCTGATACCTATTCGGGCCCCCTGATTTCTGATACA 1424
QY 440 lETTPArgGlyMetIleAla-LysGlyGlyThrProValCysCysAlaArgCysValPro 459
Db 1425 TATGAGAGAGATGATGCTCAAAAGGTGGAATCTCCGCTGTGTGCTCGATGTGATCT 1484
QY 460 MetGlyLys-GlyIleGluTTPTrpLysLeu-ProGluValaAsnCysSerAlaArgThrA 479
Db 1485 ATGGGAAGGGGAGATGAACTAAACTGCTCCAGAGTGTCAATTTGGTCCGACAGAACTG 1544
QY 479 sPLeuAsnMetLeuAlaLysHisSerTyrAlaValAla-LIleGlyCysGluIleValaPhePhe 498
Db 1545 ATTGAATATGCTGCTTAAACATTACCGCGTTGCCATGGAGATGATGATGTTTTC 1604
QY 499 ValProAspArgGluGluAspPheAlaSerTyrThrGluPheLeuArgTyrLeuSerSer 518
Db 1605 GTACCAAGCAGAGAAAGATTTTGCCTTACACTGA-TTTCTTCGCTACTTTAGCTCA 1663
QY 519 LysAspArgAlaGlyValAlaAlaLysLeuAspAspGlyThrThrLeuPheLeuValProPro 538
Db 1664 AAAGATCGGGG-GGTGTTGCCAAATTAAGATGATGATCAACTTATTTCTGGTGCCTCA 1722
QY 539 SerAspPheLeuThrAspValaLeuGlnValThrArgGlnGluArgLeuTyrGlyValaVal 558
Db 1723 TCAATTTCTAACTGATGATGATCAAGTGAACCGCTCAAGAGCGGTATATATGGTGT-GTT 1781
QY 559 LeuLysLeuProProProAlaValaProValThrAlaSerTyrThrArgGlnGluSerGlnSer 578
Db 1782 CTCAAGTTACCCCGCAGCGGTTCTGTTCAGACATCATACAGACAAAGATCTCACTCC 1841

```


QY 639 LeuThrProGluLeuLeuAlaThrLeuAlaSerIleLeuPro---AlaThrSerGlnPro 657
 |||||
 Db 352 TTAACCCCGAGCTTTAGTACTCTGCGCTTTTCTCCCTGACGCTTCTTCAACT 293
 |||||
 QY 658 AlaAlaProGluSerHISglnProMetSerGlyProSerThrValAlaSerThrAlaHis 677
 |||||
 Db 292 ACTGCTCCTGAGATGACCAAACTATGTCAGTTCCTTCATCGGTTACTTCCAGAGTACT 233
 |||||
 QY 678 GlnSerAnglyLeuTyraAnglyGluAla---ProSerGlnAlaTrpIlyAlaGlyPro 636
 |||||
 Db 232 CAC-----TACATGAGAGAACACAGCTTCTCAAGCTTGAAATGAGATCG 185
 |||||
 QY 697 GlnThrValHisAspAlaSerAnglnSerPheGlnGlnTyrglyAnglnTythrPro 716
 |||||
 Db 184 -----CAAAATTATGGGAATCGATCAATCCA 158
 |||||
 QY 717 AlaGlyGln---LeuProProProProSerArgTyrrProAlaSerAsn-----Asn 733
 |||||
 Db 157 GCTGGACACACACCTCTCTCTCTCTCCGCTTACCTCGGCTTCGAACAATCCAA 98
 |||||
 QY 734 ProAsnTyrrHrSerGlyMetValHisGlyAsnMetGlnTyrglnSerGlnSerValAsn 753
 |||||
 Db 97 CCACTACTCTCCAGTGAATGCTCATGAAACATGACATCAAGGCCAAATCTGTTAAC 38
 |||||
 QY 754 MetProGlnLeuSerProLeuProAsnMetProHisAsnAsnTyrrSer 769
 |||||
 Db 37 ATGCTCTAG-----ATGATCATATATATATACAC 8
 |||||

RESULT 3 866 bp DNA linear GSS 01-APR-2004
 CL511524
 LOCUS
 DEFINITION
 SAHL_849_F10.v1, genomic survey sequence.

ACCESSION
 CL511524
 VERSION
 CL511524.1 GI:46008844

SOURCE
 Arabidopsis thaliana (thale cress)

KEYWORDS
 Arabidopsis thaliana
 Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
 AUTHORS
 Sessions,A., Burke,B., Presting,G., Aux,G., McEliver,J., Patton,D.,
 Dietrich,B., Ho,P., Baccaden,J., Ko,C., Clarke,D., Cotton,D.,
 Bullis,D., Snell,J., Miguel,T., Hutcheson,D., Kimmery,B.,
 Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
 A high-throughput Arabidopsis reverse genetics system
 Plant Cell 14 (12), 2985-2994 (2002)

JOURNAL
 MEDLINE
 PUBMED
 12468722

COMMENT
 Contact: Sessions A

Applied Trait Genetics
 Syngenta Biotechnology Inc.
 3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
 Email: allen.sessions@syngenta.com
 ABRC Stock Number CS837982; T-DNA left border flanking sequences of
 Syngenta Arabidopsis Insertion Library (SAIL) lines are available
 through the Arabidopsis Biological Resource Center (ABRC).
 Sequences represent a pool of amplified genomic regions and not
 single contiguous sequences.
 Class: TDNA tagged.

FEATURES
 source
 Location/Qualifiers

1. 866
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /ecotype="Columbia"
 /db_xref="taxon:3702"
 /clone_lib="SAIL_849_F10.v1"
 /note="T-DNA left border sequences were isolated using a
 modified TAIL-PCR strategy"

ORIGIN

Alignment Scores:
 Pred. No.: 1,636-54 Length: 866
 Score: 808.50 Matches: 170
 Percent Similarity: 79.82% Conservative: 12
 Best Local Similarity: 74.56% Mismatches: 32
 Query Match: 16.84% Indels: 14
 DB: 9 Gaps: 3

US-09-920-705-3 (1-901) x CL511524 (1-866)

QY 488 AlaValAlaIleGlyCysGluIleValPhePheValProAspArgGluGluAspPheAla 507
 |||||
 Db 183 TCGGTGGCATTGAGATGATGATCGTTTTCGTAACGACAGGGAAGAAATTTGGG 242
 |||||
 QY 508 SerTyrrHrGlnPheLeuArgTyrrLeuSerSerTyrrAspArgAlaGlyValAlaIleu 527
 |||||
 Db 243 TCTTACACTGAATTTCTCCGTAAGCTTACTGCTCAAAAGATCGGGGGGTTCCAAATTA 302
 |||||
 QY 528 AspArgGlyThrThrLeuPheLeuValProProSerAspPheLeuThrAspValIleuGln 547
 |||||
 Db 303 GAGAGTGTACAACTTATTTGTTGCTTCATCAATCTTTCTTACTGATGATCTCAA 362
 |||||
 QY 548 ValThrArgGlnGluArgLeuTyrrGlyValValIleuTyrrProProAlaValPro 567
 |||||
 Db 363 GTGACCCCGTCAAGAACGGTATATGTTGTTTCTCAAGTTACCCCGCCAGCTTCT 422
 |||||
 QY 568 ValThrAlaSerTyrrArgGlnGlnSerGlnSerAsnProLeuHisTyrrMetAspGlnAla 587
 |||||
 Db 423 GTTACAGTATCATACAGACAGAAATCTCAATCTCTGCAATTATATGATCAAGC 482
 |||||
 QY 588 ArgAspSerProAlaAsnAlaSerHisSerLeuTyrrProArgGluAsnTyrrLeuArg 607
 |||||
 Db 483 CCGGATTCACCTGCCAATGCTATGTCAGATTTATATCTCTGAGGAAATTAATCTTAAG 542
 |||||
 QY 608 GlyAlaProGluHisLeuThrAlaAlaSerTyrrProSerValSerGluProLeuArgIle 627
 |||||
 Db 543 GGTGACCAAGAACATTTGACAGCTGCTTCAAAACATCTGTTAGGACCTCTCATATA 602
 |||||
 QY 628 ProAsnAsnAlaAlaProGlnAlaGlyValSerLeuThrProGluLeuLeuAlaThrLeu 647
 |||||
 Db 603 CCAATATATGATCCCTCCCAAGCTGGGTATGTTTAACTCCGAGCTTTTATCCACTCTG 662
 |||||
 QY 648 AlaSerIleLeuProAlaThr---SerGlnProAlaAlaProGluSerHisGlnProMetSe 667
 |||||
 Db 663 GATATATTTCTC-----TCTCTCATTAACACAAAGATACAGGCTTTTACACCTTGAA 716
 |||||
 QY 667 rGlyProSerThrValValAlaSerThrAlaHisGlnSerAnglyLeuTyraAnglyGluAl 687
 |||||
 Db 717 ACCTCCTCGGGGCGCTCTATGTTNNCTTACAAACAAAGAAAT-----TCGTCCAGGG 767
 |||||
 QY 687 aProSerGlnAlaTrpIly-----ArgGlyProGlnThrVa 699
 |||||
 Db 768 GCCGGCTCAAGNGAGAACGCCCGGTTCTCAGCTTGAGACAGAGGGCTCCACACCTGT 827
 |||||
 QY 699 HisAspAlaSerAnglnSer 706
 |||||
 Db 828 TCATTGACCGGTAAATCATCC 849
 |||||

RESULT 4

BH552335 724 bp DNA linear GSS 14-DEC-2001
 LOCUS
 BOGH093TR BOGH Brassica oleracea genomic clone BOGH093, genomic
 survey sequence.

ACCESSION
 BH552335
 VERSION
 BH552335.1 GI:17804115

KEYWORDS
 SOURCE
 ORGANISM

Brassica oleracea
 Brassica oleracea
 Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE
 AUTHORS
 Town,C.D., Van Aken,S., Uterback,T., Koo,H. and Fraser,C.M.

TITLE Whole genome shotgun sequencing of *Brassica oleracea*
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOGH093TF
Contact: Chris Town

712R
712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@ligr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
source location/Qualifiers

1..724
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOGH093"
/note="Vector: PHOS1; Site 1: BexXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BexXI linkers"

ORIGIN

Alignment Scores:
Pred. No.: 5 22e-50 Length: 724
Score: 751.00 Matches: 150
Percent Similarity: 70.29% Conservative: 18
Best Local Similarity: 62.76% Mismatches: 24
Query Match: 15.64% Indels: 47
DB: 8 Gaps: 4

US-09-920-705-3 (1-901) x BH552335 (1-724)

QY 173 GlnTPrAlaGlySerTyrAspAsnArgAsnGlyAsn---MetAsnHisLysProGln--- 190
Db 55 AAATGGCTGGCTCTTAAGATTAAGGAATGGCAATGATGATCAAAACCTCAGGTT 114
QY 190 ----- 190
Db 115 CGTTTAAATATTCAAAAAGAGCTTGTATTGTTCTGCTTATATGTATGATGATT 174
QY 191 -----TyrProHisSerTyrGluAspPheLysGlyAs 201
Db 175 AATTAGCTTTTGAAGTTTGTGTATCTTCACTCACTCAATGAAGACCCACAGAGAGA 234
QY 201 PValGlnProSerLysValLeuTrpIleGlyPheProProThrAlaThrGlnCysAsnAs 221
Db 235 TGACCAAGCAAGTAAAGTTCTGTGATGGGTACCTCCTTCTGTTCAG-----ATAGA 288
QY 221 PGIuGlnIleLeuHisAsnAlaMetIleLeuPheGlyGluIleGluArgValLysSerTyr 241
Db 289 CGAGCAATGCTCAACAAGCAATGATATCTTGGTGAGATCAAGAGAAAGTAAAGTTA 348
QY 241 rProSerAixAspPheAlaLeuValGluPheArgSerAlaGluGluAlaArgGlnCysLys 261
Db 349 CCGGTGAGGAGCATTTTCACTGTGAGTTTGAAGGCTGAGCAAGCTGCCAAGCCAA 408
QY 261 eGIuGlyLeuGlnGlyArgLeuPheAsnAsnProArgIleLysIleMetTyrSerAsnAs 281
Db 409 GGAAGGCTCAACAAGGAGGCTATTCAAGGATCCAGATCACTTCACTGATCTGAGAGA 466
QY 281 PGIuLeuProProGluGlnAspAspThrSerPheTyrSerGlyMetLysArgSerArgTh 301
Db 469 TGAATATCCTCTTCAAGCAAGTGAATAGTTTATTCTGCGCTGAACGGTCAAGGCC 528
QY 301 rAspMetPheAsnAsnAspProSerCysValSerSerProHisSerThrGlyIleProGlu 321
Db 529 AGGTATATTCATCACTATCTCT----- 550
QY 321 ySerMetArgProLeuArgGlyThrAsnGluArgSerTyrAsnGlyValaGluTyrAsnAs 341
Db 551 -TCAATTAGGCATTCAGAGGTACAGATGACCGTTCTTATATATGTTTCAAGATTACATGA 609

QY 341 PValAlaGlyLysGluProAsnTrpArgArgProSerAlaAsnGlyThrGlyIleLeuPr 361
Db 610 CGTTGTTGTATGAGCAAACTGAGAGGCGCTCTCCAAATGAGCATGGGATATCTCC 669
QY 361 oSerProThrGlyProGlyIleLeuProSerProAlaGlnGlyThrArgArgPro 379
Db 670 ATCTCCAGCAGCAGCATGGATCTCCCTCCTCTCTGCAAGAGATAGGAACCT 724

RESULT 5
BX840244
LOCUS BX840244
DEFINITION BX840244 Arabidopsis thaliana Hormone Treated Callus Col-0
ACCESSION Arabidopsis thaliana cDNA clone GSUTPH87ZF04 SRIM, mRNA sequence.
VERSION BX840244.1 GI:42534327
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS Castelli,V., Aury,J.M., Jallion,O., Wincker,P., Clepet,C., Menard,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V., Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
TITLE Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences:
A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation
Unpublished (2004)
Contact: Genoscope

JOURNAL
COMMENT Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr

The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli V., Aury J.M., Jallion O., Wincker P., Menard M., Cruaud C., Schachter V., Weissenbach J., Salanoubat M.
UNRG INRA : Clepet C., Caboche M.
Annotation is based on the June 2003 version of the Arabidopsis genome released by MIPS (Munich Information center for Protein Sequences).
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/EST
http://www.genoscope.cns.fr/cgi-bin/g9b/g9b?source=Arabidopsis.

FEATURES

source location/Qualifiers
1..812
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="GSUTPH87ZF04"
/tissue_type="Hormone Treated Callus"
/clone_lib="Arabidopsis thaliana Hormone Treated Callus Col-0"

ORIGIN

Alignment Scores:
Pred. No.: 1.3e-49 Length: 812
Score: 747.00 Matches: 148
Percent Similarity: 86.84% Conservative: 17
Best Local Similarity: 77.89% Mismatches: 23
Query Match: 15.56% Indels: 2
DB: 5 Gaps: 0

US-09-920-705-3 (1-901) x BX840244 (1-812)

QY 1 MetAlaLeuSerMetLysProPheArgAlaAspAspSerGlyPheGlnSerAsnLeu 20
Db 246 ATGGCTTATCTATGAAGCATTCAGAGCGGATTCGGTTCACAGTCAAAACAATCTT 305
QY 21 TrpValGlySerLeuThrProGluThrThrGlnLysSerAspLeuThrGluLeuPheGlyArg 40

Db 306 TGGGTCGGTAGCCTTAACCGCCGAGACGACAGTCAAGTTCGACCGAGTTTGGCAGA 365
 Qy 41 TGTGTAAPPIleAspArgIleThrValIysSerSerArgIlyPheAlaPheIleTyrTyr 60
 Db 366 TAAAGCGAAATGATAGATGACCTCCGGTATATCTTCACGAGGCTTGGCTTATCTCTAC 425
 Qy 61 ArghIsvAlGluGluAlaValAlaAlaIysGluAlaLeuGlnGlyAlaAsnLeuAlaGly 80
 Db 426 AGACATGTGAGGAGCGAGCTGCGACCCCAAGAGGCTCTTCAGAGAGCAAAATTTAAATGCG 485
 Qy 81 SerGlnIleIyIleGluTyrAlaArgProAlaIysProCysIysSerIleTyrValGly 100
 Db 486 AGTCAAAATTCAGATAGATAGCAAGACCGCCCAAGCTTGTCTCAGATCTATGGGTGGT 545
 Qy 101 GilyIleGlyProAsnValSerIysAspAspLeuGluGluGlnPheSerIlyPheGlyIys 120
 Db 546 GCGATCGGCACTAATGATCCAAAGATGACAGACCAAGTTCACCAAGTTTGGGAAA 605
 Qy 121 IlegluAspPhe-ArgPheLeuArgGluArgIyIleThrAlaPheIleAspTyr-TyrGlu 140
 Db 606 GTCAAGATTTTACGGGTTCTCTCAGAAACGACAGATTTCTTATATTCATAGAGA 665
 Qy 140 etAspAspAlaLeuGluAlaIysSerIleAsnGlyIysProMetGlyIyIleSerPheLeu 160
 Db 666 TGGATGATGATTTACCAAGATAGAGCATGACTGGGAGACCATGGGTGCTGCTTTTGC 725
 Qy 160 rGValAspPheLeuArgSerGlnAlaProIyIysGluGlnIleTyrAlaGlySerTyrAsp 180
 Db 726 GTGTTAAATTTCTCCGACACCAAGACAAAGAGAGAAATGGGCTGACTTTCAGATA 785
 Qy 180 snArgAsnGlyAsnMetAsnIleIys 188
 Db 786 ACAGAAATGGAATAGATATATATA 811
 RESULT 6
 CDS0A2CE
 LOCUS 2706 bp mRNA linear HTC 04-FEB-2004
 DEFINITION Arabidopsis thaliana Full-length cDNA Complete sequence from clone
 GSLTFB492D02 of flowers and buds of strain col-0 of Arabidopsis
 thaliana (thale cress).
 ACCESSION BX842306
 VERSION BX842306.1 GI:42406938
 KEYWORDS
 SOURCE HTC; GSLT cDNA.
 ORGANISM Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 2706)
 Castell, V., Aury, J.M., Jallion, O., Wincker, P., Clepet, C.,
 Menard, G., Cruaud, C., Quetier, F., Scarpelli, C., Schachter, V.,
 Temple, G., Caboche, M., Weissenbach, J. and Salanoubat, M.
 Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences:
 A Combined Approach to Evaluate and Improve Arabidopsis Genome
 Annotation
 Unpublished
 2 (bases 1 to 2706)
 Genoscope.
 Direct Submission
 Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 The sequences are based on single pass reads.
 Life Technologies (a division of Invitrogen) members carried out
 full-length libraries construction : Temple G.
 Genoscope members carried out sequencing and annotation : Castell
 V., Aury J.M., Jallion O., Wincker P., Menard M., Cruaud C.,
 Schachter V., Weissenbach J., Salanoubat M.
 URGV INRA : Clepet C., Caboche M.
 Annotation is based on the June 2003 version of the Arabidopsis
 genome released by MIPS (Munich Information center for Protein
 Sequences). 5 prime and 3 prime are assembled with Phrap.

http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
 length
 http://www.genoscope.cns.fr/cgi-bin/g9b/g9b?source=Arabidopsis.
 FEATURES
 Location/Qualifiers
 source
 1. 2706
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Col-0"
 /db_xref="taxon:3702"
 /clone="GSLTFB492D02"
 /issue_type="flowers and buds"
 /plasmid="pCMVSPORT_6"
 1. 2706
 /gene="At4g12630-At4g12640"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,16e-46 Length: 2706
 Score: 716.50 Matches: 271
 Percent Similarity: 38.738 Conservative: 112
 Best Local Similarity: 27.408 Mismatches: 298
 Query Match: 14.924 Indels: 310
 DB: 3 Gaps: 37
 US-09-920-705-3 (1-901) x CDS0A2CE (1-2706)
 Qy 17 SerAsnAsnLeuTyrValGlySerIleThrProGlnThrGlnSerAspLeuThrGlu 36
 Db 108 TCGCAGACATCTTGGGTGGTGGTGAACCTTCCTCAGGATATCTTGAACGAAATTCAGCAT 167
 Qy 37 LeuPheGlyArgTyrGlyIysAspIleAspArgIleThrValIysSerSerArgGlyPheAla 56
 Db 168 CGGTTCTTAAGATTTGGTGAATTTGGAAGCTTAGGCTTTCAGCTTGAAGAGCATACGCA 227
 Qy 57 PheIleTyrTyrArgIleValGluGluAlaValAlaIysGluAlaLeuGlnGlyAla 76
 Db 228 TTTGTTAAATTTTAAACAGATGAGATGCGTTCGCTGATGATGATGCTTCAAGGCTTC 287
 Qy 77 AsnLeuAsnGlySerGlnIleIleGlyIleGluTyrAlaArgProAlaIysProCysIysSe 96
 Db 288 CCACCTCTGGAACCCACTTAGATCGAGTTTGTCTAAGCGCGGAG----- 336
 Qy 96 rLeuTyrValGlyGlyIleGlyProAsnValSerIysAspAspLeuGluGlnPheSe 116
 Db 337 -----TC 338
 Qy 116 rlyPheGlyIyIleGluAspPheArgPheLeuArgIyIleThrAlaPheIleAs 136
 Db 339 ATCAACTGATCAGCGCACTGAT-----GA 362
 Qy 136 pTyrTyrGluMetAspAspAlaLeuGlnAlaIysSerMetAsnGlyIysProMetGlyIyl 156
 Db 363 TATATATGCCCATGAA-----AAACAACGGCTTAAAMAAAAAGAGGTC 407
 Qy 156 ySerPheLeuArgValAspPheLeuArgSerGlnAlaProIyIysGluGlnTyrAlaGly 176
 Db 408 GTCTTTGTCCAA----- 420
 Qy 176 ySerTyrAspAsnArgAsnGlyAsnMetAsnIleIysProGlnTyrProHisSerTyrGly 196
 Db 421 -----AGGAGCTCCAGAAATGCGTTAT-----GAGAGCCCAACACATATAG 461
 Qy 196 uAspPheIys-----GlyAspValGlnProSerTyrValLeuTyrIleGlyPhePr 213
 Db 462 CAATCAAAATGATATAGATAGAAATGCAAGAACCGAGAGGATCTATATATAGGCTTCC 521
 Qy 213 oProThrAlaThrGlnCysAsnAspGluGlnIleLeuHisAsnAlaMetIleLeuPheGly 233
 Db 522 T-----GCTTCGTTGAAGATGATGAGCTCTTGAAGAGAGCTTTCTTCGTTGG 575
 Qy 233 yGluIleGluArgValIysSerTyrProSerArgAsnPheAlaLeuValGluPheArgSe 253
 Db 576 AGAAATACAAAGGTCACGGTATTCCTGCTGCGCATTTATGCAATTCGAAA 635

OY		253	TALGIUCLUNLAAAGGInCySlySGULUGLYLeuGLINclLyArgLEUPheAsnAenProAr	273
Dd		636	TCTGTATGGACCTCTGTATAAGCGGAAGAAGTCTTCAGGAAAATTATTGGCAATCCTCG	695
OY		273	GILEySIleMeTYrSerAAsnAPGlu-----	282
Dd		696	AGTGATATTGTTGTTTGCMAAGAGGMACTTCCCATCTGGCACTGSAAGGGTCCGTC	755
OY		283	-----LeuProGluGln-----AspSPhtSerPheTYrSerGlyme	296
Dd		756	GGGTGATCATCTCTCCACCCTTCACAGATCTGTGTGATGACTGGGATCTTCAGAAAGTTTA	815
OY		296	FLyASrSerArqThr---AspMet.PheaAsnAsnAProSerCyValIserSerProHl	315
Dd		816	TCTTCAGATACAAACTATGGAAGCATACAGAAATTCCTCACT--GTAGAGAACCAACA	872
OY		315	S-----	315
Dd		873	TTACATAGAGATAGAGATTAGAAAGATTCTGAAGTTACATTTTTCACAGAAACGAGA	932
OY		316	-----SerThrGlyIleProGlySerMetArGProlEuMrGGLyThraEngLuARgSe	333
Dd		933	CTCAAGTACCATGAGAGGTCTCCGATATGAAAGGTGCG--AGATCTACGCATTAAGATTCC	989
OY		333	RTyrAnGILyAlaGluTYrAsnAP-ValIValGlyLySGluProAsnTPRArgArpROS	353
Dd		990	TCAGGATATGCAATGAAATATCATGTAAGTCTGGGGAAA-----TGGGATCTTCAATT	1040
OY		353	eRIla-----AnGlyThrGlyIleLeuProSerProThrGlyProGlyIleL	369
Dd		1041	CCGTGATATCTCACAGATTCCAGACAGAGATT---CAGAGATGAAAGCCATGGGATCT	1097
OY		369	eUProSerProAlaGlnGlyThrArg---ArgPrometArGPserAnProAspSerT	387
Dd		1098	ACCGAAATGACTCTACTACACAGAAATCAAGAGATTGAAGACAGATCTCGCACCC	1157
OY		387	rPGluGlyTYrAspProIa--GlnLeu-----ValArgLuserLyAsrThrArgA	404
Dd		1158	TGAGAGCAGCTC-CCAGGGCATCAGCTTTCTGTATAGACAGAAAGACGCCCTTTT	1216
OY		404	rGaBpGlySerValAspGlyPheThrPrometGlyValAspGluARgSerPheGlyArgG	424
Dd		1217	CMAAGGATCTGCTGAT---TTTTCCCAAAGATGCAATTGAGCGAATCATGAGCTG	1273
OY		424	IY-----	424
Dd		1274	GACAGCTAAGATACAAACCAACGATTGAGCAGCCCTTAACCTGCTTAAGAAATGGGG	1333
OY		425	-----Serv	426
Dd		1334	ACAAAGTAGCTTACAGAACCAACATGAAAGATTAAAGGAGTTATCTTCGCCATCAG	1393
OY		426	aLIAlaIARgProIIeArGlyProProABP-----SerAspHisI	440
Dd		1394	TTGTCCCTGAAGAAAAAGTACACCCCTGATCATGATCAGCATCACTAAAGATCGGA	1453
OY		440	IeTPARgLyMeTIleAlaIaySGlyGlyThrProvalCySCynAlaARGCyValProm	460
Dd		1454	ACTGGGAAGGAGCATATTGCTTAAGGAGGAATCCCATTTGTTCAGACAAATAGCTTCTCG	1513
OY		460	eGlyLySGlyIleGluThrLySLeuProGluValAlaSnCySerLAlarGThrAspL	480
Dd		1514	TGGGCAAAAGTAGGACATGATGCTGCCGTGAAGTTTCTAGATTGACGGCAAGACTGGT	1573
OY		480	eAsnMeCLeuAlaIySHISryAlaValAlaleGlyCySGluIleValPhePheValP	500
Dd		1574	TAGACATGCTGCTAAGCATTAATCTACAAATATCAATAACGATGGGTGTTTCTTTGTC	1633
OY		500	roASPArgLUglunAPhealaserTyrrhGluPheLeuArgTyrlEusSerSerlyBA	520
Dd		1634	CTGGAAGATGCTCTATATCGTGTTTTATGATGATTATTAATCATTAATCTCGAGAGGAAGC	1693

QY	520	sparglaaglyvalalaaleuleuhsparaglyuthrthleuPheuleuvalproprosera	540
Db	1694	AACGGGACGCTGTTTAAATTTGATGACAAACAACGCTGTTCTGGAGCCCTCATCTG	1755
QY	540	srpPheuleuthrAspValleuGlnValThirArgGlnIaArgleutyrglyValValleu	560
Db	1754	ATTTCCTCCGAAGAAAGTACTTAAAGTTCCTGGGAAACTAAGCATCTCTGGAGTTATCTAC	1811
QY	560	ysleuProProProAlaValProValThralaserTyArgGlnIuSerGlnSerAsn	580
Db	1814	GTTTA-----	1811
QY	580	roleuHieTyMetserGlnIalargAserProAlaAsnIaserHiserleutyP	600
Db	1819	-----GAATGGCTGTTCTGGCTCCGGTGGAC	1844
QY	600	roProArgIuAsnTyrlleArgGlyAlaProGlnHieLeuThralaIaserIySpros	620
Db	1847	CTGTTTCACACGACGAGTGAAGAAAGATACAGACTTCTAATCTTAATGTGTAAACAT	1900
QY	620	eValSerGlnProleuArgIleProAsnAsnIalalaProGlnIaGlyValSerleut	640
Db	1907	CATATTCAGAA-----	1911
QY	640	hrProGlnIuLeuIalThrlleuIaserIleleuProAlathSerGlnProAlaIal	660
Db	1918	-----GTAAGTGGAGCGTTTCCCTGA--TGNGAAMACCCGATATTC	1954
QY	660	roGluSerHieGlnProMetSerGlyProserThnValIaserThralHieGlnserA	680
Db	1957	CG-----GGTCCAAACGCTTTCTGGAGTAgTGGGGCTGTGACA	1997
QY	680	snGlyLeuTyAsnGlyGluIalProserGlnIalalTryIuAsGlyProGlnThralH	700
Db	1996	AT-----CAAGCATCATAGATCATATC-----GTGC	2022
QY	700	IsAspAlaSerAsnGlnIserPheGlnGlnTyrglyAsnGlnTyThrProAlaGlyGln	720
Db	2026	AGGACAAACATGATGATGATGATGATCACCGGATTTCTGGA-----	2063
QY	720	eupProProProProSerArgTyrrProProAlaSerAsnAsnProAsnTyThr	737
Db	2064	-----TCGGACTGGCGCGCTCGTATACAAACCCGAAAGTTCAACCTTCA	2104
QY	738	-----SerGlyMetValHieGlyAsnMetGlnTyrglnSerg	750
Db	2110	TAGATCATACTGTTCAAAAACAACGCGGTTTGTCTCGGGAACAACGAA-----	2155
QY	750	InserValAsnMetProGlnIuSerProleuProAsnMetProHieAsnAsnTySerH	770
Db	2160	-----AATCCAGATCTTAAGTC-----	2177
QY	770	eTyTyThrGlnGlySerSerAsnHieSProValSerGlnProMetValGlnGlnTyrglnP	790
Db	2176	GTTAT-----CATGATACGGAATCTCCAGTCCCAACAGAGGTTTTCAGC	2211
QY	790	roGln-----AlaSerMetProAsnGlnAsnTyrglyProIlePros	804
Db	2218	CTGAACAGCTAACACATTAGCTCCCTCGCTACCAACAGAC-----C	2255
QY	804	eTyTyGlnGlnIalAsnPheHieGlyValThrThrasnGlnIalAsn-----L	821
Db	2260	AACAACAGGAACAGAAC-----ACGCCAATTCAGCTGAGAGATTTTTTTTTT	2301
QY	821	euaAsnProSerGlnPheGln-----AlaAlaMetGlnProProAlaAspIyAlaAsn	839
Db	2308	TAAAGCCACAGTGTGTTTTCAGCATTTGGCAATGACAGACAGACGCAACATCTCAACCG	2367
QY	839	eugIuProGlnIalAsnGlnIalAleuArgleuGlnProMetIleSerGlyAsnGlyGlnTyT	859
Db	2368	TAACTCCCAAAATCAAAATGATGACAGATACA-----AGTGCACACGCCCAACAC	2411
QY	859	hrThAspGlyGlyValAspIyAsnGlnIaGlyTyrglnSerThrIuGlnInPheAlaIal	879


```

Db      2419 AAGAAAGACAGAGCCCATCCACAGAGCCTCTTCAGACATTCAGCTACGACCTG 2478
Qy      879 snleuleuGlnTleGln 886
Db      2479 GACTTCTCCAGCATTCACAA 2501

RESULT 7
BG543956          602 bp      mRNA      linear      EST 01-MAY-2002
LOCUS            E1706 Chinese cabbage etiolated seedling library Brassica rapa
DEFINITION       subsp. pekinensis cDNA clone E1706, mRNA sequence.
ACCESSION        BG543956
VERSION          BG543956.1 GI:20374936
KEYWORDS         EST.
SOURCE           Brassica rapa subsp. pekinensis
ORGANISM         Brassica rapa subsp. pekinensis
                 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE        1. (bases 1 to 602)
AUTHORS          Ryu,S.H., Kang,J.S., Kang,C.-h., Kim,C.Y., Choi,Y.J., Lee,S.-H.,
                 Bahk,J.D., Lee,S.Y., Cho,M.U. and Lim,C.O.
TITLE            Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
JOURNAL          Unpublished (2001)
COMMENT          Contact: Lim, C.O.
                 Plant Molecular Biology & Biotechnology Research Centre
                 Geongsang National University
                 #800 Gazea-dong, Jinju 660-701, Korea
                 Tel: 82 55 751 6255
                 Fax: 82 55 759 9363
                 Email: colim@nongae.gsnu.ac.kr
                 Seq primer: T7.

FEATURES
source           1..602
                 location/qualifiers
                 /organism="Brassica rapa subsp. pekinensis"
                 /mol_type="mRNA"
                 /cultivar="Jangwon"
                 /sub_species="pekinensis"
                 /db_xref="taxon:51351"
                 /clone="E1706"
                 /tissue_type="Etiolated seedling"
                 /lab_host="Xu-1 Blue"
                 /clone_1b="Chinese cabbage etiolated seedling library"
                 /note="Vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"

ORIGIN
Alignment Scores:
Pred. No.:      1.15e-45      Length:      602
Score:          695.50        Matches:    138
Percent Similarity: 84.02%    Conservative: 25
Best Local Similarity: 71.13% Mismatches:    26
Query Match:    14.49%       Indels:      6
DB:             4            Gaps:       2

US-09-920-705-3 (1-901) x BG543956 (1-602)

Qy      23 GlySerIeuThrProGluThrGluSerAspLeuThrGluLeuPheGlyArgTyrGly 42
Db      13 GGCAGCTGACCCCGACTACTCGAATCGATCGCGATCTGTTCGGAAGTTCCGCG 72
Qy      43 AspIleAspArgIleThrValTyrSerSerArgGlyPheAlaPheIleTyrTyrArgHis 62
Db      73 GATATGACAGGCTCACCGCTATCTTCCGTTGCTTCGCTTATCTCACTTCAGGCGC 132
Qy      63 ValGluGluAlaValAlaAlaTyrGluAlaLeuGlnGlyAlaAsnLeuAsnGlySerGln 82
Db      133 GTTGAAGAACCTTAAGCCCGGAGAGGCTCTTCAGAGGTCACATCTAAACGAGGTCCG 192
Qy      83 IleTyrIleGluTyrAlaArgProAlaTyrProCysTyrSerLeuTyrValGlyGlyIle 102
Db      193 ATTAAATGCAATTCGACGACCGGCAAAACCTGTGAAGTCTTGGGTTGTGGATTC 252

```

```

Qy      103 GlyProAsnValSerIysAspIleuGluGluPheSerIysPheGlyIysIleGlu 122
Db      253 AGCTCAGTGTCCCGAAGGATGATCGAGGAGGATTCAGAGATTTGGGGAATCGAA 312
Qy      123 AspPheArgPheLeuArgGluArgIysThrAlaPheIleAspTyrTyrGluMetAspArg 142
Db      313 AGTTTCAGGTTCTTAAAGATCGCAAGACGCTCTTTATTGACTATTTTGAACGTGACGAT 372
Qy      143 AlaLeuGlnAlaLysSerMetAsnGlyIysProMetGlyGlySerPheLeuArgValAsp 162
Db      373 GCTGTACAGCAAAAGCATGAAACGGGAAGCCAGTAAAGTGGTATTTTGGAGTTGAT 432
Qy      163 PheLeuArgSerGlnAlaLysProLysIysGluGlnTTPAlaGlySerTyrAspAsnArgAsn 182
Db      433 TTTCTTGCTCACAGGCGCAAGAAAGAACCAACGCGCTTGCTCCGAAGATGCGAGGAT 492
Qy      183 GlyAsnMetAsnHisIys---ProGlnTyrProHisSerTyrGluAspPheIysGlyAsp 201
Db      493 GGCAGTTGATAGTCAGACACAGCATATCTACTCA-----CTTAGAGAT 540
Qy      202 ValGlnProSerIysValLeuTTPleGlyPheProPorphr 215
Db      541 GGCCAAACA-AGCAACGTTCTGTGATGGGTGACCTCCTTCT 581

RESULT 8
BE037904          701 bp      mRNA      linear      EST 07-JUN-2000
LOCUS            A06B09 AA Arabidopsis thaliana cDNA 5', mRNA sequence.
DEFINITION       BE037904
ACCESSION        BE037904
VERSION          BE037904.1 GI:8332920
KEYWORDS         EST.
SOURCE           Arabidopsis thaliana (chale cress)
ORGANISM         Arabidopsis thaliana
                 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE        1. (bases 1 to 701)
AUTHORS          Bohner,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M.,
                 Ferreira,H., Kawasaki,S., McCollough,A., Michalowski,C.B.,
                 Palacio,C., Scara,G., Wheeler,M. and Zepeda,G.R.
TITLE            Functional Genomics of Plant Stress Tolerance
JOURNAL          Unpublished (2000)
COMMENT          Contact: Michalowski,C.B.
                 University of Arizona
                 Bio Sciences West room 513, Tucson, AZ 85721, USA
                 Tel: 520-621-7982
                 Fax: 520-621-1697
                 Email: chm@u.arizona.edu
                 Best blastx match: 'gb|AA64314.1| (AC002335) hypothetical protein
                 [Arabidopsis thal. . . 165 3e-40'. An open reading frame exists.
                 Insert Length: 2 Std Error: 0.00.

FEATURES
source           1..701
                 location/qualifiers
                 /organism="Arabidopsis thaliana"
                 /mol_type="mRNA"
                 /ecotype="Columbia"
                 /db_xref="taxon:3702"
                 /tissue_type="leaves, flowering plants"
                 /dev_stage="12 weeks"
                 /clone_1b="AA"
                 /note="20 h 200mM NaCl"

ORIGIN
Alignment Scores:
Pred. No.:      2.28e-45      Length:      701
Score:          693.00        Matches:    140
Percent Similarity: 96.62%    Conservative: 5
Best Local Similarity: 94.59% Mismatches:    3
Query Match:    14.43%       Indels:      2
DB:             2            Gaps:       0

US-09-920-705-3 (1-901) x BE037904 (1-701)

```

Qy 1 MetalaleusertlethProphetaArglaAspaspSerglyPheGlnSerAsnAnleu 20
 Db 245 ATGGCGTTATCTATGAAACCATTCAGAGCCCATGATTCGGTTCCAGTCAAAACATCTT 304
 Qy 21 TrpValGlySerleuthrProGluThrThrGlnSerAspleuthrGluLeuPheGlyArg 40
 Db 305 TGGGTGGGTAGCCTTAACCCGAGAGCAGACAGATCAGATCTGACCGAATGTTGGAGA 364
 Qy 41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgIlyPheAlaPheIleTyr 60
 Db 365 TACGGCGATATTGATGATACACGGTGTATCTTCACAGAGCTTGGCTTTATATACACTAC 424
 Qy 61 ArgHisValGluGluAlaValAlaAlaValGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
 Db 425 AGACATGTGAGAGAGACGATCGACCCAAAGAGCTCTTCAAGAGCAATTTGAATGGA 484
 Qy 81 SerGlnIlelySileGluTyrAlaArgProAlaIlyProCysIlySerleuthrValGly 100
 Db 485 AGTCAAAATTAAGATCAATACGACAGACCGCAAAACCTTGTAAAGATCTATGGGTGGGT 544
 Qy 101 GlyIleGlyProAsnValSerIlyAspAspleuGluGluGluPheSerIlyPheGlyLys 120
 Db 545 GGAATCGGTCTTAATGTTCCAGAGATGACCTGACAGAGAGATTCAAGACCTTGGGAAA 604
 Qy 121 IleGluAspPheArgPheLeuArgGluArgLysThrAlaPheIleAspTyrTyrGluMet 140
 Db 605 ATCAGGATTTTAAAGTCTTCAAGACACGACAGACGACGCTTTCATGCA-TATTATGAGATG 663
 Qy 141 AspAspAlaLeuGlnAlaIlySer 148
 Db 664 GCTGATGC-TTACAGGCTTAAGACA 686

RESULT 9

LOCUS AU237346 635 bp mRNA linear EST 01-APR-2002
 DEFINITION AU237346 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-16-A15 5',
 mRNA sequence.
 ACCESSION AU237346
 VERSION AU237346.1 GI:19876515
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 635)
 Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
 Akiyama, K., Enju, A., Oono, Y., Sakurai, T., Carninci, P., Kawai, J.,
 Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A.,
 Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.
 Large scale analysis of Arabidopsis full-length cDNA
 Unpublished (2002)
 Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: meeki@rtc.riken.go.jp
 An Arabidopsis full-length cDNA library was constructed essentially
 as reported previously (Seki et al., 1998) cDNA cleaved with BamHI
 and XhoI was ligated to modified Lambda FL-C1 vector (Carninci et
 al., submitted for publication) digested with BamHI and SalI. This
 clone is in a modified pBluescript vector. Please visit our web
 site (http://www.gsc.riken.go.jp/e/Plant/index_e.html) for further
 details.

FEATURES

SOURCE
 1..635
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAFL16-16-A15"
 /lab_host="DH10B"

ORIGIN
 /clone lib="RAFL16"
 /note="Site_1: BamHI; Site_2: SalI; dark-grown"

Alignment Scores:
 Pred. No.: 1.61e-38 Length: 635
 Score: 607.00 Matches: 123
 Percent Similarity: 97.62% Conservative: 0
 Best Local Similarity: 97.62% Mismatches: 2
 Query Match: 12.64% Indels: 0
 DB: 1 Gaps: 0
 US-09-920-705-3 (1-901) x AU237346 (1-635)

Qy 1 MetalaleusertlethProphetaArglaAspaspSerglyPheGlnSerAsnAnleu 20
 Db 255 ATGGCGTTATCTATGAAACCATTCAGAGCCCATGATTCGGTTCCAGTCAAAACATCTT 314
 Qy 21 TrpValGlySerleuthrProGluThrThrGlnSerAspleuthrGluLeuPheGlyArg 40
 Db 315 TGGGTGGGTAGCCTTAACCCGAGAGCAGACAGATCAGATCTGACCGAATGTTGGAGA 374
 Qy 41 TyrGlyAspIleAspArgIleThrValTyrSerSerArgIlyPheAlaPheIleTyr 60
 Db 375 TACGGCGATATTGATGATACACGGTGTATCTTCACAGAGCTTGGCTTTATATACACTAC 434
 Qy 61 ArgHisValGluGluAlaValAlaAlaValGluAlaLeuGlnGlyAlaAsnLeuAsnGly 80
 Db 435 AGACATGTGAGAGAGACGATCGACCCAAAGAGCTCTTCAAGAGCAAAATTTGAATGGA 494
 Qy 81 SerGlnIlelySileGluTyrAlaArgProAlaIlyProCysIlySerleuthrValGly 100
 Db 495 AGTCAAAATTAAGATCAATACGACAGACCGCAAAACCTTGTAAAGATCTATGGGTGGGT 554
 Qy 100 YGlyIleGlyProAsnValSerIlyAspAspleuGluGluGluPheSerIlyPheGlyLys 120
 Db 555 TGGATCGGGCTTAATGTTCCAGAGATGGGTGGAGAGAGATTGCAAGATTGGGAAA 614
 Qy 120 eIleGluAspPheArg 125
 Db 615 ATC-GAGGATTTTAAAG 629

RESULT 10

LOCUS BH851246 416 bp DNA linear GSS 13-JUN-2002
 DEFINITION SALK_072730.56.00.x Arabidopsis thaliana TDNA insertion lines
 Arabidopsis thaliana genomic clone SALK_072730.56.00.x, genomic
 survey sequence.
 ACCESSION BH851246
 VERSION BH851246.1 GI:21422117
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 416)
 Alonso, J.M., Leisner, T.J., Barajas, P., Chen, H., Cheuk, R.,
 Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,
 Shinn, P., Zimmerman, J. and Becker, J.R.
 A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome
 Unpublished (2001)
 Contact: Joseph R. Becker
 Salk Institute Genomic Analysis Laboratory (SIGNAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of
 TDNA. This sequence lies within an annotated exon of At2g34340.
 Class: TDNA tagged.

TITLE
 JOURNAL
 COMMENT

FEATURES

Location/Qualifiers
1. 416
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="SALK_072730.56.00.x"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/cdna_protocols.html"

ORIGIN

Alignment Scores:

Pred. No.: 4,3e-35 Length: 416
Score: 561.00 Matches: 107
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.69% Indels: 0
DB: 8 Gaps: 0

US-09-920-705-3 (1-901) x BH851246 (1-416)

QY 795 ProAnglnAsnTyrGlyProIleProSerTyrGlnGlnAlaAsnPhelSGIValThr 814
DB 2 CCNAACCAAACTATGCTCATTCCTCAAGTATGAGAGCTATATTTCAATGGCGTACA 61
QY 815 ThrAnglnAlaGlnAsnLeuAsnProSerGlnPhelGlnAlaMetGlnProProAla 834
DB 62 ACAATATGAGGACAGAACTTAAACCTTCCCAATTTCAACTGCGACGACCAACGACGA 121
QY 835 AspValAlaAsnLeuGlnProGlnAsnGlnAlaLeuArgLeuGlnProMetIleSerGly 854
DB 122 GATAAGGCAATTTTAAAGCCACAAACCAAGCACTACAGATTCAGCTTAAATCTCTGGG 181
QY 855 AspGlyGlnGlyThrThrAspGlyGluValAspValAsnGlnArgTyrGlnSerThrLeu 874
DB 182 GATGCTGACGGTACACGATGGGAGCTGATAGATCAGAGATCCAGTCAACCTA 241
QY 875 GlnPheAlaAlaAsnLeuLeuGlnIleGlnGlnIleGlnGlnGlnGlnGlnGlnGln 894
DB 242 CAATTTCACACAAACCTTCTTCCAGATACAGAGAAACAGCAGCAACGCTTCAAGGT 301
QY 895 ThrProAlaGlyGlnGlyPro 901
DB 302 ACTCGGCTGACACGGGGCT 322

RESULT 11
AV829561 582 bp mRNA linear EST 01-APR-2002
LOCUS AV829561
DEFINITION Arabidopsis thaliana cDNA clone RAFL09-47-P10 5',
mRNA sequence.
ACCESSION AV829561
VERSION AV829561.1 GI:19871621
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidops.
1 (bases 1 to 582)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Contact: Motoaki Seki
COMMENT Unpublished (2002)
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center

FEATURES

Location/Qualifiers
1. 582
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL09-47-P10"
/dev_stage="plants at various developmental stages from germination to mature seeds"
/lab_host="DH10B"
/clone_lib="RAFL09"
/note="Site 1: BamHI; Site 2: SalI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"

ORIGIN

Alignment Scores:

Pred. No.: 5.99e-33 Length: 582
Score: 537.00 Matches: 108
Percent Similarity: 97.30% Conservative: 1
Best Local Similarity: 97.30% Mismatches: 2
Query Match: 11.19% Indels: 0
DB: 1 Gaps: 0

US-09-920-705-3 (1-901) x AV829561 (1-582)

QY 1 MetAlaLeuSerMetIleProPheArgAlaAspSerGlyPheGlnSerAsnLeu 20
DB 254 ATGGCTTATCTATGAGCATTCAGAGCCATGATTCGGTTCCAGTCAACAACTTT 313
QY 21 TrpValGlySerLeuThrProGlnThrThrGlnSerAspLeuThrGluLeuPheGlyArg 40
DB 314 TGGGTCGATGACCTTAACCGCGAGACGACAGATCAGATCGACGATGTTGTTGAAGA 373
QY 41 TyrGlyAspIleAspArgGlyLeuValTyrSerSerArgGlyPheAlaPheLeuTyr 60
DB 374 TACGGGATATGATGAAATCAGCGTGTATCTTACAGAGCTTTGCTTATATATAC 433
QY 61 ArgHsValGlnGluAlaValAlaAlaLysGlnAlaLeuGlnGlyAlaAsnLeuAsnGly 80
DB 434 AGACATGTGGAGGAAGCAGTCCGACCAAGAGCTCTTCAAGAGCAAAATTTGAATGA 493
QY 81 SerGlnIleLysIleGlnTyrAlaArgProAlaLys-ProCysLysSerLeuTrp-ValG 100
DB 494 AGTCAAAATTAAAGTACGACGACGACGACGACGACGACGACGACGACGACGACG 553
QY 100 LysGlyIleGlyProAsnValSerLysAsp 109
DB 554 GTGGAATCGCCCTATATGTTCCAAAGAT 582

RESULT 12
CO095058 817 bp mRNA linear EST 16-JUN-2004
LOCUS GR_Ea17103.r GR_Ea Gossypium raimondii cDNA clone GR_Ea17103 3',
mRNA sequence.
ACCESSION CO095058
VERSION CO095058.1 GI:48793744
KEYWORDS EST.
SOURCE Gossypium raimondii
ORGANISM Gossypium raimondii

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.


```

Db 207 GATTGATATGCTGGAAAAAACAACCTAAGAGAGCTTCGGGTTTGAAATGCTTCTTC 266
Qy 499 ValProAspArgLgluGluAspPheAlaSerThrGluPheLeuValGlyLeuSerSer 518
Db 267 TTACCAAGACAGTGAAGATGATTTGTTGCTTACCTGAATTTCTGCGGTACTGGGGCTCA 326
Qy 519 LysAspArgLgluValAlaValLeuAspAspGlyThrThrLeuPheLeuValProPro 538
Db 327 AAAGTGGCGCAGGGGTGTTAAGTTGAAGCAAGAAACCACTTTGTTTGGTCCGCCCA 386
Qy 539 SerAspPheLeuThrAspValLeuGlnValThrArgGlnLysLeuValGlyValVal 558
Db 387 TCGGATTTCTGACGAGATGTTCTGCAAGTGTGACAGCCCTTATGCGGTGTA 446
Qy 559 LeuLysLeuProPro-----ProAlaValProValThrAlaSerThrArgGln 575
Db 447 CTGCACATCTCCACAGATGCTCAATGCTGCTCCAGAGACCTGTACTGACAGCAGAG 506
Qy 576 SerGlnSerAsnProLeuHisThrMetAspGlnAlaArgAspSerProAlaAsnAlaSer 595
Db 507 TCACAAAGCT-----TATATGAT-----GGTGT 530
Qy 596 HisSerLeuThrProProArgLysAsnThrLeuArgGlyAlaPro-----GluHis 612
Db 531 GATACCACTGCTAGCAGATGCAAGAAATTAACATATGAGCTCTGCTAATGCGCAACCATCAT 590
Qy 613 LeuThrAlaAlaSerLysProSerValSerGlu----- 623
Db 591 CAGGATCTGATTAATCGGGGCTCTTTCGTGAGAGACAGTTCAGTCAAGGTATCAAGC 650
Qy 624 ---ProLeuArgLysProAsnAlaAlaProGlnAlaGlyValSerLeuThrProGln 642
Db 651 TTCCCACTG-----ACCAAAATGACAGACAGACAGACAGTCAACAGCCGAT 701
Qy 643 LeuLeuAlaThrLeuAlaSerLysLeuProAlaThrSerGlnProAlaAlaProGlnSer 662
Db 702 ATTATGCGCACTTATGCGCAAACTTATGCGCA---AATGCGATCATATGTTCCAGCAAAAT 758
Qy 663 HisGlnProMetSerGlyProSerThrValValSerThrAlaHisGlnSerAsnGlyLeu 682
Db 759 AGTCAG-----GTGGGTAATCTTACAGCAATCCCGTCAGCAG 794
Qy 683 TyrAsnGlyLysLysProSerGln-----AlaTyrLysArgGlyPro 696
Db 795 TTGCGCAGGCAAGCTCCACACAGCTATGCGGCGCATGTTGCTGACAGACAGCAGCA 854
Qy 697 GlnThrValHisAspAlaSerAsnGlnSerPheGlnGlnThrGlyAsnGlnThrPro 716
Db 855 -----ATACATCTGCT-----TACAAACCTGAGGTTACCTTA 887
Qy 717 -AlaGlyLysLeuProProProSerArgThrProProAlaSerAsnAsnProAsnThr 736
Db 888 AGCT-----TTGCTCTCCACAAGCT-----CCTCTCCCTTAAAAAGAACCCNGT 932
Qy 736 ThrThrSerGlyMetValHisGlyAsnMetGlnThrGlnSer----- 749
Db 933 T-----TCTCTCCCTTACACAGCAGATTTTAAACAGAAAACCTT 971
Qy 750 -----GlnSerValAsnMetProGlnLeuSerProLeuProAsnMetProHisAs 766
Db 972 TTACCCCTCTCAAAAAATTAAACAGTCCCTTATGTAACAG--TTAACCAAGCTTA----- 1021
Qy 766 AsnThrSerMetThrThrGlnLysSerSerAsnHisProValSerGlnProMetValG 786
Db 1022 -----ATTGCCGCTCCAAAAAATATCA 1041
Qy 786 nGlnThrGlnProGlnAlaSerMetPro---AsnGlnAsnThrGlyPro 801
Db 1042 ACCTTACCGGGCC-----CTCCCAACCAAAAAAATGATGCGCC 1081

```

```

LOCUS CF322873 582 bp mRNA linear EST 18-AUG-2003
DEFINITION HDN--02-E23.g1 OshDACL-overexpressing transgenic rice lambda phage
CDNA library II (HDN) Oryza sativa (japonica cultivar-group) cDNA
ACCESSION CF322873
VERSION HDN--02-E23, mRNA sequence.
KEYWORDS CF322873.1 GI:33793966
SOURCE EST.
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 582)
AUTHORS Kim,J.S., Jun,K.M., Cheong,P.J., Kim,M.J., Lee,T.H., Shin,Y.C.,
Song,S.I., Kim,J.K., Kim,Y.-K. and Nahm,B.H.
TITLE Large-scale Sequencing Analysis of Rice ESTs
JOURNAL Unpublished (2003)
COMMENT Contact: Nahm B.H.
Genomics and Genetics Institute, Greengene Biotech Inc., Division
of Bioscience and Bioinformatics, Myongji University
Yongin, Kyonggi, Korea
Tel: 82 31 330 6193
Fax: 82 31 321 6355
Email: bhnahm@gbio.com, bhnahm@bio.myongji.ac.kr.
FEATURES
source
location/Qualifiers
1..582
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nackdong"
/db_xref="taxon:39947"
/clone="HDN--02-E23"
/cisue_type="callus"
/dev_stage="proliferated callus on 2N6 media for 2 weeks"
/lab_host="E.coli SOLR"
/clone_lib="OshDACL-overexpressing transgenic rice lambda
phage cDNA library II (HDN)"
/notes="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:
XhoI; cDNA was inserted into lambda uni-ZAP XR vector at
5' end with EcoRI and 3' end with XhoI site. mRNA was
derived from rice Histone Deacetylase overexpression
line."

```

ORIGIN

```

Alignment Scores:
Pred. No.: 6,71e-31 Length: 582
Score: 511.50 Matches: 107
Percent Similarity: 68.21% Conservative: 26
Best Local Similarity: 54.87% Mismatches: 49
Query Match: 10.65% Indels: 13
DB: 7 Gaps: 5

```

US-09-920-705-3 (1-901) x CF322873 (1-582)

```

Qy 438 AspHisIleThrArgLysMetIleAlaValGlyThrProValCysCysAlaArgCys 457
Db 8 GAGCACTTCTGGGTGAGTACATTGCCAAGAGTGAGATCTCTTAATGTCGCTCGCTGT 67
Qy 458 ValProMetGlyLysGlyLysGlnThrLysLeuProGlnValValAsnCysSerAlaArg 477
Db 68 TTCCCTATTAAGAGAGGAGGTGTGAGATACCTTTACCGGAGTGTGTTATGTTCCGCTAGA 127
Qy 478 ThrAspLeuAsnMetLeuAlaValHisThrValAlaValAlaIleGlyCysGlnIleValPhe 497
Db 128 ACTGAGCTGATGATGCTGGCAAGCACTTATAGAGACGCTTCAGAGGTTGATATCGCTTC 187
Qy 498 PheValProAspArgLgluGluAspPheAlaSerThrThrGluPheLeuValGlyLeuSer 517
Db 188 TTCTTGCAGATAGTGAAGATGACTTTGTTCTTACACTGAATTTTTCGCTACTTAGGC 247
Qy 518 SerLysAspArgLgluValAlaValLeuAspAspGlyThrThrLeuPheLeuValPro 537
Db 248 TCAAAAGTCGGCAGGGGTTGTAAGGTTGATGCGGAGCTATTTATTTTGTGCCCA 307

```

```

QY 538 ProSerAspPheLeuThrAspValIleuGlnValThrArgGlnIleuArgLeuTyrGlyVal 557
Db 308 CTTTCCGATTTTTCAGAAATGTTTGCAAGTTGATGCTCCAGACGCTTTACGGGTGA 367
QY 558 ValIeuLysLeuPro-----ProProAlaValProValThrAlaSer 571
Db 368 GTATTACACATTCGCAAAATGTCTGCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTGTA 427
QY 572 TyrArgGlnIle-----SerGlnSerAsnProLeuHisTyrMetLaspGlnAlaArg 588
Db 428 CAAAGGCCACACTTAACCTGCACACCAATCAAACT-----TACTACGATGAA--AGG 478
QY 589 AspSerProAlaAsnAlaSerHisSerLeuTyrProProArgGlnAsnTyrIleArgGly 608
Db 479 GAAATTCCTTTGCGAGAGAGATATATATATACGCTTACCAACATCCACATCGAGAT 538
QY 609 AlaProGlnHisLeuThrAlaAlaSerLysProSerValSerGln 623
Db 539 GCT---GATCACCGTGGATCTTTACGTGAGGATTCATTCATCATCAA 580

```

Search completed: April 8, 2005, 13:44:55
 Job time : 7010 secs

and (pc) m7.